



人工智能

算法岗

江湖武林秘籍(中)

献给各路豪杰



算法岗位面试必备

现有一本

江湖之中风云莫测

武林盟

算法岗位面试必备



人工智能算法岗江湖目录

武林秘籍（上册）

武林秘籍练功必读	第 1 页
1. 字节跳动算法岗武功秘籍.....	第 6 页
2. 阿里巴巴算法岗武功秘籍.....	第 68 页
3. 腾讯算法岗武功秘籍.....	第 104 页
4. 百度算法岗武功秘籍.....	第 137 页
5. 华为算法岗武功秘籍.....	第 180 页
6. 美团算法岗武功秘籍.....	第 205 页
7. 京东算法岗武功秘籍.....	第 229 页
8. 网易算法岗武功秘籍.....	第 256 页

武林秘籍（中册）

9. 拼多多算法岗武功秘籍.....	第 276 页
10. Vivo 算法岗武功秘籍.....	第 298 页
11. 招银网络算法岗武功秘籍.....	第 311 页
12. 360 公司算法岗武功秘籍.....	第 326 页
13. 海康威视算法岗武功秘籍.....	第 339 页
14. 快手算法岗武功秘籍.....	第 354 页
15. 小米算法岗武功秘籍.....	第 375 页
16. 深信服算法岗武功秘籍.....	第 392 页
17. 作业帮算法岗武功秘籍.....	第 405 页
18. 滴滴算法岗武功秘籍.....	第 420 页
19. 蚂蚁金服算法岗武功秘籍.....	第 439 页
20. 顺丰科技算法岗武功秘籍.....	第 453 页
21. 依图科技算法岗武功秘籍.....	第 464 页
22. 旷视科技算法岗武功秘籍.....	第 476 页

23. 微软算法岗武功秘籍.....	第 494 页
24. Oppo 算法岗武功秘籍.....	第 507 页
25. Bigo 算法岗武功秘籍.....	第 518 页
26. 猿辅导算法岗武功秘籍.....	第 534 页

武林秘籍（下册）

27. 中兴算法岗武功秘籍.....	第 546 页
28. 商汤科技算法岗武功秘籍.....	第 557 页
29. CVTE 算法岗武功秘籍.....	第 573 页
30. 大华算法岗武功秘籍.....	第 583 页
31. 欢聚集团算法岗武功秘籍.....	第 600 页
32. 平安科技算法岗武功秘籍.....	第 610 页
33. 大疆算法岗武功秘籍.....	第 622 页
34. 蘑菇街算法岗武功秘籍.....	第 631 页
35. 云丛科技算法岗武功秘籍.....	第 642 页
36. 追一科技算法岗武功秘籍.....	第 653 页
37. 爱奇艺算法岗武功秘籍.....	第 663 页
38. 搜狗算法岗武功秘籍.....	第 674 页
39. 地平线机器人算法岗武功秘籍.....	第 683 页
40. 58 集团算法岗武功秘籍.....	第 693 页
41. Keep 算法岗武功秘籍.....	第 704 页
42. 寒武纪算法岗武功秘籍.....	第 712 页
43. 搜狐算法岗武功秘籍.....	第 724 页
44. 有赞算法岗武功秘籍.....	第 733 页
45. 知乎算法岗武功秘籍.....	第 742 页

9|拼多多算法岗武功秘籍

1 拼多多面经汇总资料

第一节
拼多多面经
汇总资料

(整理: 江大白)
www.jiangdabai.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 拼多多面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 拼多多面经-87 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 全职岗位类

【NLP 工程师】、【图像视觉算法工程师】、【搜索广告算法工程师】、【数据分析工程师】、【核心搜索组搜索算法工程师】、【拼越计划算法工程师】

1.3 面试流程时间安排

拼多多面试流程整理：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	根据项目深挖
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	根据项目深挖
第三面	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：

- 要不是技术+技术+HR，要不就是技术+HR+技术

1.4 拼多多面试心得汇总

★ 两轮技术面试都是，先撕代码再聊项目，手写的相对较快，面试反馈比较正向，后面给了SSP。

★ 面试的时候聊了很多实习期间的东西，基本上整体回顾了一下自己的经历，复杂网络、推荐系统等东西都聊了很多，比较重视深度和广度，细节也问的比较多。

★ 面试前的最佳复习资料还是自己的简历，简历上写的东西一定要会讲原理会推导。

最后，不要在简历上写你不能熟练说出大部分细节点的知识！面试官基本还是看你简历提问多的。

★ 问项目的时候，一定要显示出自己对项目细节了如指掌，并且熟悉背后的原理，这样就算项目很水，如果能【显得自己基础牢靠】，面试官也会满意的。

★ 总体来说 ML 基础问题比较简单，算法题比较难（但不强求），主要问项目。

★ 拼多多感觉主要考察点就是从项目开始发散，所以项目里用到的东西一定要认真准备，当然没有项目可能连笔试机会也没，所以项目，论文，实习怎么也得有一个的。然后编程题主要

就是简单题为主，但是需要会简单题的优化算法。

★ 面试前就是几本常用的机器学习系列书籍，还有 svm，lr 这一类的推导，认真复习，自己想好对于自己的项目可能会问到什么，提前准备好，还是很有用的。

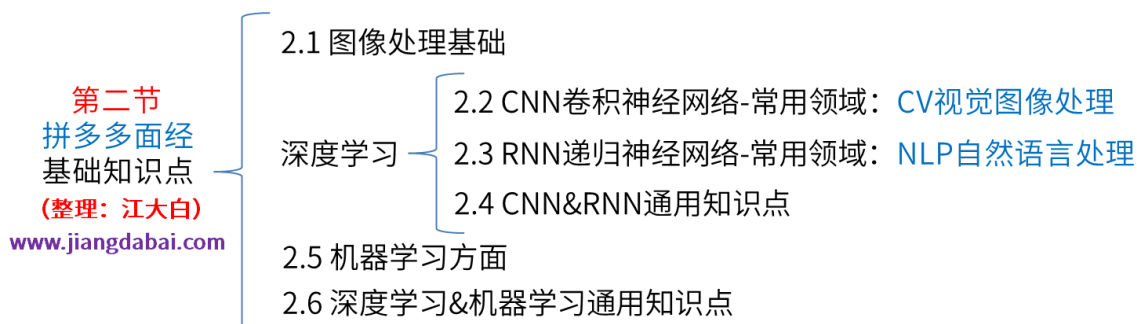
★ 总结：一面总体偏基础，于简历展开询问，要对简历内容熟稔于心，基础要牢固，同时也要有一定的深度广度。二面开始强调项目或者文章内容。

★ 感觉电商的场景还是比较单一的，主要是刀具等违禁品检测识别。

拼多多分工比较明确，这点应该是个优势，可以增加算法团队的专注度。

★ 整体而言题量不大，但是问的相当细致（尤其是 SVM 和 transformer 那块）。难度也不大，考察的知识点都集中在常规的模型和算法上面。

2 拼多多面经涉及基础知识点



2.1 图像处理基础

无

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 卷积方面

- 卷积如何加速？
- 卷积和反卷积的原理说一下

- 讲一下 Dropout 层的原理？
- Dropout 的 train 和 test 的区别？
- Dropout 会改变数据分布，导致训练和测试样本分布不一致，怎么解决这个问题？

2.2.1.2 池化方面

- 池化和反池化的原理说一下？

2.2.1.3 网络结构方面

- Resnet 结构是怎么样的，有什么优点？
- Vgg 结构的优点是？
- inception 结构的优点？
- SeNet 的结构用过吗？
- Densenet 结构讲一下？

2.2.1.4 其他方面

- 数据增强的一些方法？
- 梯度消失和梯度爆炸是什么原因，怎么避免或解决？
- 讲一下 BN 层的原理？
- 什么是 roi pooling，怎么实现的？
- 什么是 roi align，怎么实现的？

2.2.2 数学计算

- 求普通卷积和 resnet block 的参数量、计算量，并对比两种结构？

2.2.3 公式推导

- 解释一下 CNN 原理，卷积后 feature map 尺寸大小怎么算，写一下公式？

2.2.4 手写算法代码

- 实现一个卷积操作，可以用 numpy?

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- RNN 的改进有哪些？讲 LSTM 和 GRU，对于更长的序列怎么处理？
- 说说你理解的 LSTM（RNN 到 LSTM 讲 LSTM 的缺点，引出 transformer）？
- LSTM 如何调参？
- RNN 怎么解决，LSTM 为什么可以解决？

2.3.2 手绘网络原理

- 画出 LSTM 的结构图？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 你认为 self attention 在提一些什么模式的特征呢？
- self-attention 和普通 seq2seq 的 attention 区别？

2.4.2 模型评价

- Recall 和 Precision 的原理讲一下？
- 用过哪些衡量分类性能的指标？
- 用了哪些评价指标，AUC 是什么？
- 写一个函数计算 AUC？

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

- 推一下 PCA 的公式？
- SVD 了解吗？主要的作用是什么？F 范数下的最优近似？

② 特征选择

- 特征如何选择，贡献性？共线性？
- 单特征选择有哪些方式？组合特征呢？

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）

- GDBT 和随机森林的区别？
- 随机森林是 bagging 还是 boosting？

A.基于 bagging：随机森林

- 随机森林如何选择特征，如何评价特征重要性？
- RF 说一下，RF 树特别多会发生什么，RF 和 GDBT 哪个深一些，为什么？
- 随机森林的随机性在哪儿？

B.基于 boosting：Adaboost、GDBT、XGBoost

- XGBoost 和 GDBT 的区别，提升在哪里？
- 讲讲 XGBoost 原理？ XGBoost 优势？ XGboost 怎么解决过拟合？过拟合了，怎么调参数？

- GBDT 关键参数有什么，树深和树的棵数都会导致过拟合，如果现在模型过拟合了，这两个参数调哪个？
- GBDT 是怎么选择特征的，答借助于 CART 树模型进行选择，类似于 ID3,C4,5 用信息增益和信息增益率？GBDT 还可以构建特征什么的？
- GBDT 树怎么生成的，残差从哪里来的，实际用的时候用的负梯度，绝对值损失函数的时候怎么求导？
- XGBoost 的损失函数是什么？用过哪个损失函数？
- 项目中使用了 lightgbm，让讲一下 gbdtd？

② K 近邻 (KNN)

- KNN 是生成式还是判别式，时间复杂度，怎样优化？
- KNN 最后做回归，除了取平均还有什么其他方法？

③ 逻辑回归 LR

- 逻辑回归的介绍，如何训练，几种训练方法？
- LR 的定义，目标函数，优化方法？
- LR，随机森林，XGBoost 区别，细节？
- LR 和决策树的区别是什么？
- LR 损失函数：怎么用最大似然求损失函数？
- 说到过拟合，讲讲 LR 怎么应对过拟合？(L1, L2) 为什么正则化可以？
- 逻辑回归如何分类出非线性超平面？

④ SVM (支持向量机)

- SVM 原理讲一下？
- 通俗介绍 SVM？
- 讲一下 SMO 算法？
- SVM 的损失函数 (hinge)
- SVM 的推导-对偶--求解？核函数的选择？

- 为什么 SVM 这么有效？核函数能把特征映射到高维空间？
- 一般怎么选核函数？SVM 跟 AdaBoost 有什么联系？
- LR 和 SVM 介绍+区别，什么场景用 SVM 比较好？
- SVM 和神经网络的区别，两个分别适用于什么场景，SVM 有过拟合问题吗？
- 拉格朗日乘子，KKT 原理？

⑤ 朴素贝叶斯 (Naive Bayes)

- 贝叶斯和传统统计学派区别吗？那从贝叶斯定理的角度说为啥 L2 可以解决过拟合？
- 知道共轭分布吗？

⑥ 决策树 (DT)

- 决策树有几种 (cart/id3/c4.5)，他们分别有什么特点？
- 3 种决策树，区别，以及一个连续特征如何建树？
- 各种决策树模型的优劣（从最简单的 ID3 到最后的 LGB）？
- 树模型（决策树）复杂度？
- ID3 复杂度是多少，怎么优化呢？
- 决策树了解吗，GBDT 怎么做的，做分类问题用的什么树？

2.5.1.4 无监督学习-聚类方面

- 介绍一下 kmeans 的原理？
- 看简历上用过 GMM，讲解 GMM 原理？

2.5.2 手推算法及代码

- 推导 SVM？
- 推导 LR？
- GMM、EM 算法手推？
- 推导一下 GBDT 的公式？
- 手推 SVM+KKT 条件+SMO 算法简单描述？

- 推了一下 LR 的损失函数以及梯度传递

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

- 交叉熵的公式？
- MSE 的公式？
- 问逻辑回归算法的原理？
- 逻辑回归推导一下，损失函数+求解过程
- 为什么不用回归的 loss？均方误差？
- MSE 是不是凸函数，是 MSE 本身的问题吗？
- 逻辑回归表达式是不是凸函数？

2.6.2 激活函数方面

- 激活函数有哪些？Relu 和 Sigmoid 的区别？

2.6.3 网络优化梯度下降方面

- 项目中的超参是怎么调的，比如学习率或者优化方法，都用过哪些方法？
- adagrad 用过没？跟 SGD 有何区别？
- SGD 里面的“S”，什么意思？
- 训练里面用过多卡吧？里面对梯度有啥操作？
- 为什么要对多卡的梯度做平均？
- 为什么梯度下降可以求解最优化，数学理解？
- 深度学习里面优化器，ADAM 特点？
- 网络权重初始化方法、优化方法 Adam、多卡训练时超参数怎么变化（lr、迭代次数）

2.6.4 正则化方面

- 有哪些正则化、L1 和 L2、L1 梯度，在零点怎么办？

- L2, L1 原理(答：高斯先验，拉普拉斯先验)
- L1 正则化损失函数如何求解？
- L2 正则化的特点，使用场景？
- L1 和 L2 正则化的区别，特点？

2.6.5 过拟合&欠拟合方面

- 深度学习中怎么应对过拟合？
- 如何解决过拟合？过拟合、欠拟合、如何调参？

2.6.6 其他方面

- 深度学习模型和机器学习模型的区别是什么？
- 数据不平衡怎么做？

3 拼多多面经涉及项目知识点

第三节
拼多多面经
项目知识点
(整理：江大白)
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

3.1.1.1 讲解原理

- 项目用到了 faster rcnn，然后就扯了 RCNN 系列，RCNN 和 CNN 的区别？
- rcnn 到 faster rcnn，介绍一下，然后具体每一步操作都问的很细，roi pooling 具体怎么做的、RPN 是怎么做的，输入输出是什么？
- Faster rcnn 的最后输出大小是多少、它的正负样本是怎么选择的？

- NMS 具体怎么做的，假设这是一个函数，那么这个函数输入输出是什么，中间操作又是怎么做的？
- 问我改个 6 点怎么从 Faster rcnn 改到 yolov3?
 - (1) backbone 要变
 - (2) ROI pooling 要删掉
 - (3) anchor 的设置要变
 - (4) 正样本的设置方式要变
 - (5) regression loss 的实现要变
 - (6) yolov3 用了多尺度训练的方式
- 问目标检测目前的难点和未来的发展方向？
- Faster RCNN 中 RPN 的细节

3.1.1.2 手写代码

手写 SoftNMS 代码

手写 NMS 代码

3.2 深度学习：RNN 递归神经网络方面

3.2.3 自然语言处理 NLP

① Bert

- 说一下 bert 几个最关键的点？
- 问了下 bert 的输入以及和 gpt, elmo 的区别？

② Transformer

- 介绍 transformer 结构，手推计算讲解细节，画 Transform 结构图？
- 讲 transformer 如何并行化运算？
- 说一下 transformer 的优点？

③ CRF

- LSTM+CRF 层中 CRF 的作用?

④ HMM 隐马尔科夫模型

- HMM 和 CRF 的区别?

⑤ Word2vec

- Word2vec 原理?讲一下 word2vec 怎么实现?
- Word2vec 和 doc2vec 区别? 和 fasttext 的区别?
- Word2vec 里有什么重要的方法, 负采样和层次 softmax, 分别是怎么实现的, 为什么能提升速度和效率?
- 怎么用 word2vec 训练 embedding 的, 怎么做的? 召回分哪几路? 具体候选集是多少?
- Word2vec 怎么训练的, 为什么近义词能训练出距离比较小的向量?
- LDA 的词表示和 word2vec 的词表示有什么区别?

⑥ CNN 方面

- TEXTCNN 和 TEXTRNN 比较, fasttext 原理说明, 为啥比其他效果差?

⑦ fasttext (词向量和文本分类)

- Fasttext 原理, 为什么用 skipgram 不用 cbow, 负采样怎么做到, 公式是什么?

⑧ 其他

- EM 算法, 为什么隐变量的问题要用 EM 算法?
- 怎么做文本分类, RNN 和 CNN 各自特点?

3.3 强化学习

- 说一下 GAN 的原理, 有什么应用, 有什么变种?

3.4 机器学习方面

3.4.1 推荐系统

- 介绍 FM、FFM?
- DEEP FM 讲一下?
- deepfm 协同过滤算法
- 推荐和搜索有什么区别? (面试官是做搜索的)

4 数据结构与算法分析相关知识

第四节
拼多多面经
数据结构与算法分析
(整理: 江大白)
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

- 求最长上升子序列?
- 一维数组, 最长上升子序列?
- 【1,2,3,4, 7,8,8,8,9】有序数组, 给定 k, 找到 k 的最后一次出现的索引?
- 找数组中第 k 大的数
- 数组里面有一个数出现的次数超过一半, 这个数是哪个?
- 旋转有序数组的查找?
- 旋转数组查找 target 的开始和结束索引?
- 数组先升序再降序, 找到最大的数返回? 注意边界条件。

- 从[7,2,3,4,9,1,7,5,20,6]找两个数，使得后面一个数减前一个数的差最大？
- 合并两个有序数组，用了 python 的 pop 操作，让我解释原理？
- 给一个数组，输出子集
- 两个数组的公共元素
- 给出一个整数数组 a，给出一个整数数组 b（无重复元素），根据 b 的元素顺序，逐个判断 a 中的元素对应位置是否等于这个值，如果相等输出 1，不等对应元素输出 0。
- 无序数组构建平衡二叉树

4.1.1.2 链表

链表反转，我用了迭代写法

- 给个链表 1->2->3->4->5->6

把它变成 1->6->2->5->3->4

- 给定一个链表和一个数值，把小于数值的放到前面，大于的放到后面？
- 链表快排

4.1.1.3 字符串

- 反转字符串？
- 字符串编辑距离？（两个字符串 s,t，可以增删改，将 s 转化成 t 最小步数）
- 给定字符串 s，求与 s 编辑距离为 2 的字符串集合？
- 假设一个字符串中除了一个字符外每个字符都连续出现正好两次，找出那个字符？（例如 aabbccdd, accbbdd 等）先写了一遍 $O(N)$ 的解法，然后在面试官的引导下想到了 $O(\log N)$ 的解法。
- 字符串 A 交换两个字符任意次数能得到字符串 B，则 AB 是同一类，给定字符串求有多少类？
- Pattern 匹配：判断字符串 S 是否匹配 Pattern (如 'abbc')？刚开始说构建一个字典，同时遍历 S 和 P，但是会出现两个字符同时代表一个 S 中的字母，遂加了个 set 对 S 中出现过的

字符串进行记录。

● 最长回文子串

4.1.2 树

4.1.2.1 二叉树

● 找二叉树中两个节点的公共父节点，我一开始说的递归，让我想个非递归的方法，我说了个先层次遍历然后用一维数组存放，然后再用完全二叉树的方法去找。

● 字典树

● 给一个二叉树和一个目标值，找到和等于这个值的所有路径？

● 给两个字符串 s1 和 s2，将其表示为二叉树，问 s2 是否能由 s1 进行任意多次交换任意节点的左右子节点得到？s1, s2 的二叉树分查不限。

● 又问了复杂度和平均递归深度？

● 最大二叉搜索树

● 给一个二叉树的根节点，一个节点 p，一个节点 q，找出 p, q 最近的公共祖先

● 判断是否为平衡二叉树？

● 二叉树的层序遍历？

● 树的非递归遍历？

● 二叉树中序遍历，允许用递归写

● 二叉树的深度遍历？

● 二叉树的后序遍历

● 二叉树从左到右遍历叶子结点，输出从根节点到叶子节点的路径？

● 一个普通的二叉树，从里面找到一个节点数最多的二叉搜索树（子树），输出这个树的节点个数？

给定 dic=['ab','abc','abcd','bcd','bcde','bde','efg'], str='abcdefg', 求最长匹配 dic 中的元素，

比● 如本例就是输出['abcd','bcde','efg']。

实现：把 dic 的元素建成前缀树，然后搜索前缀树根节点，如果匹配就一直搜到根节点。

- 给了一棵多叉树的所有的边，让返回一个合理排序，要求子节点必须排在父节点前面？
- 数组实现二叉搜索树？

4.1.2.2 堆

- 建堆的复杂度 ($O(N)$)

4.1.3 排序

- 计算 Topk？
- 快排，时间复杂度，最坏情况复杂度，如何改进避免最坏复杂度？
- 写一个堆排序

4.1.4 搜索

- 一个二维矩阵上的 DFS 问题:给你一个二维矩阵，每个点是一个 0-1 之间的像素值，规定超过像素阈值的称之为高点击诱导点，由相连的高点击诱导点组成的区域面积大于面积阈值的称之为高点击诱导区域，统计这个二维矩阵中高点击诱导区域的个数？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 给一幅地图上色，相邻区域颜色不能相同（我用 BFS 做的）？
- 有 10 万行的数据，每一行有三个数据，分别是起始 ip 地址，终止 ip 地址，和这个 ip 段对应的区域 比如说：

[11.12.10.01, 11.12.255.255, 北京]

然后写一个函数，输入一个 ip 地址，返回对应的区域？

- 一套扑克牌 52 张，随机发 N 张牌，判断 N 张牌有没有同花顺？

4.2.2 智力题

- 一个圆被分成 M 个扇形，一共有 N 种颜色，相邻扇形不同色，一共有几种涂法？

- 给一个股票价格序列，只准买卖一次，求什么时候买入什么时候卖出能够获得最大利润？
- 有 n 堆砖头，第 i 堆砖头的重量为 W_i ，合并第 i 堆砖头和第 j 堆砖头需要耗费能量为 $W_i + W_j$ ，问合并 n 堆砖头最少需要耗费多少能量，以及合并的具体过程是什么？
- 不同长度的绳子有不同的价值，一根绳子如何切分可以让总价值最大。动态规划求解即可？
- 一个棋盘，起始点在左上，终点右下，棋盘上有一些棋子，找到一条路径，使得经过的棋子最多，并且记录下这条路径？
- 现有一组会议起止时间，会议的时间信息 $[[s1, e1], [s2, e2], \dots]$ ，总共同时需要多少会议室，例： $[[0, 30], [5, 10], [15, 20]]$, $res=2$ 。

4.3 其他方面

4.3.1 数论

- 面积期望
- 扑克牌 123456，抽到 6 停止，求期望？
- 抛硬币，正面继续抛，反面不抛。问抛的次数的期望？
- $100w$ 个桶， $100w$ 个小球，小球随机入桶，空桶的期望（二项分布）
- 袋子里面有 n 个不同的小球，每次随机从里面取一个出来。每个球被取到的概率为 $probs[i]$ ，取到观察后再放回去。每个球至少被取到一次，需要的最少次数的期望。
- 重复数字不连续排列？
- 输入正整数 N ，从 1 开始打印不超过 N 位的所有数字，比如：

$N=1$, 打印 1, 2, 3, ..., 8, 9

$N=2$, 打印 1, 2, 3, ..., 98, 99

$N=3$, 打印 1, 2, 3, ..., 998, 999

一开始以为是快速幂，本质是用字符串来存储 N 位大数，从 0 开始用字符串加法逐次加 1，然后打印出和，等最高位进位的时候就停止打印

- 数学专业的题目：

- (1) 对数几率怎么理解?
- (2) 凸函数的定义还记得吗?
- (3) x^2 是不是凸函数?
- (4) 凸函数对导数的要求?

4.3.2 概率分析

- 出了一个统计概率题，扔硬币，然后求极限那种？

概率 1: 11 个球，1 个特殊球，两个人无放回拿球，问第一个人取到特殊球的概率？

概率 2: 11 个球，1 个特殊球，两个人有放回拿球，问先拿到这个特殊球的概率？

- 求圆内接三角形过圆心的概率
- 从 n 个数据中抽取 m 个数据，保证每个数据被抽到的概率为 m/n ?
- 求出 n 个色子之和为 s 的各种概率（可以用递归和动归做）
- 10W 个人有一个人生病，检测无病的 1%假阳性，有病一定检测出，现有一个人检测有病，则真实有病的概率？

4.3.3 矩阵运算

- 二维矩阵中，重复数字组成的最大面积？
- $N \times N$ 矩阵顺时针旋转 90° ，要求时空复杂度尽可能低？

4.3.4 其他

- 为啥泰勒展开成二阶，作用？
- 已知 X 的 2 范数=1 ($n \times 1$)， a ($n \times n$)，求 aX 的 2 范数最小值（拉格朗日）
- 斐波那契数列 $O(\log(n))$: x^2
- 最高效的方式实现幂函数？
- 数轴上的最长连续线段（要求 $o(n)$ 以内）
- 集合的所有子集
- 一亿个数字，找中位数？

- 给数字 n ，依次打印 $1 \sim 10^n - 1$ (全排列) ?
- 手撕代码，实现 LRU 算法

4.4 Leetcode&剑指 offer 原题

- Leetcode 10
- Leetcode 45: 跳跃游戏
- Leetcode 233: hard 题
- Leetcode 151 题
- Leetcode 原题: 顺时针打印矩阵

5 编程高频问题: Python&C/C++方面

第五节
拼多多面经
编程高频问题
(整理: 江大白)
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

- python 字典的底层数据结构?
- for 适用的数据类型?
- 深拷贝和浅拷贝, 如何操作?
- 单引号, 双引号, 三引号的区别?
- python 中的多线程与多进程

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 内存相关

- 说一下 C++中如何防止内存泄漏？

5.2.1.2 区别比较

- 虚函数和纯虚函数有什么区别？

5.2.1.3 讲解原理

- 头文件、命名空间、传值和引用等

5.2.1.4 讲解应用

- Const function(const) const，这三个位置的 const 有什么区别？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
拼多多面经
操作系统高频问题
(整理: 江大白)
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

6.1.1 基础问题

- SQL 的考察包括：取数，排序，日期的操作，等等。
- SQL 连接方式的区别？
- 主键和索引有什么关系？
- 索引的类型？

- mysql 底层索引，怎么实现的？

6.1.2 手写代码

- 表 1: userID,age,city

表 2: userID,pv,orderNum

Q1: 城市为上海，北京，pv>1000,orderNum <10?

Q2: 在表 2 增加一个 dt（时间，精确到天），还是上述问题

Q3: 有什么优化方式？

6.2 操作系统方面

6.2.1 线程和进程相关

- 线程和进程的区别？
- 进程之间可以通过指针共享内存吗？
- 死锁的原因及解决办法？

7 技术&产品&开放性问题

7.1 技术方面

- 如何度量两个短视频的相似度？
- 情景题：计算两个短文本的相似度，你有什么方案？假设两个短文本里面的实体词无法准确抽取（没有大量的实体词库），你又要怎么修改方案？
- 视频抽帧做分类任务，有的图片比较模糊怎么办？
- 假设你有一批数据，一个新样本，怎样在这批数据中找到与它最相近的样本
- 如果让你设计一个预测商品点击率的怎么去做，常见的决策树 ID3,C4.5,CART 有什么区别？用 cart 好还是 c4.5 好？
- 场景题：拼多多在线海量以物搜物的问题

如何标注 (active learning)

如何训练 (online learning)

如何使用？不断增加新的商品怎么办？(不重新训练，直接计算高维特征)

- 10w 的物料库，对应的用户信息，给用户推荐 top10 的物料，怎么设计算法？
- 排序用的什么模型？特征是什么？具体是什么？为什么用这个模型？和其他模型的比较？

给一句话，怎么提取出里面的地名，要先分词吗？

- 问了我机器学习、深度学习是通过什么渠道自学的？我回答的是看书（西瓜书、机器学习实战、python 深度学习、数据结构与算法等）、看网课（coursera 网站上吴恩达机器学习、深度学习的网课以及斯坦福大学网课）、参加比赛（kaggle）、刷题（leetcode、牛客）

7.2 产品方面

- 场景题：给两个商品，判断是不是一个款式？
- 场景题：怎么把 CV 一些模型用到拼多多手机找同款这个任务里面？
- 场景题：假设现在有一个商品的详情页面，希望你根据商品的介绍，自动化生成一些评论，你怎么做？ 如何发现商家卖与店家不符的商品？

7.3 开放性问题

- 如何生成用户画像？

10|vivo 算法岗武功秘籍

1 vivo 面经汇总资料

第一节 vivo面经 汇总资料 (整理: 江大白) www.jiangdabai.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 vivo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: Vivo 面经-61 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【NLP 算法实习工程师】

(2) 全职岗位类

【图像算法工程师】、【语音算法工程师】、【广告推荐算法工程师】、【nlp 算法工程师】、【芯片算法工程师】、【机器学习算法工程师】、【vivo 提前批图像算法工程师】、【深度学习算法岗工程

师（互联网方向）】、【广告推荐算法工程师】

1.3 面试流程时间安排

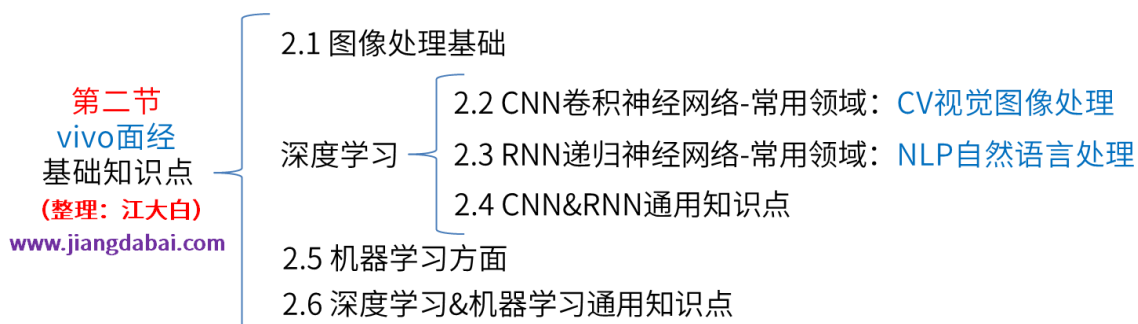
Vivo面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答	针对项目， 问改进点，创新方式
第二面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

1.4 Vivo 面试心得汇总

- ★ 建议把项目仔仔细细过一遍，涉及到的一些常见的概念，算法和原理（特别是你提到，用过的）一定要弄清楚。
- ★ 面试官对于面试者之前做的东西挺重视的，毕竟能从中问你很多问题。所以对自己做过的东西一定要很熟悉，很清楚。如果你做的东西不是很多，那做过的那些一定要连细节都很清楚。
- ★ 总体感觉，不是很难，基本第一面都是基于项目来问，都是比较基础的问题。

2 vivo 面经涉及基础知识点



2.1 图像处理基础

2.1.1 讲解相关原理

- 知道哪些噪声，图像降噪的原理？分别怎么处理？
- 知道哪些传统的降噪算法？哪种滤波器不会影响图像边缘？
- 高斯滤波器？
- 边缘检测算子，canny 梯度幅值如何计算？
- 彩色如何变灰度？

2.1.2 手写算法代码

- 写傅里叶变换？

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 卷积方面

- 卷积的物理意义是什么？
- 图像感受野怎么结算？感受野的意义与公式？
- Dropout 的工作机制，Dropout 在训练过程如何使用？

2.2.1.2 网络结构方面

无

2.2.1.3 其他方面

- 问了下常用的深度学习模型，以及框架最重要的组成部分？
- 简单叙述卷积神经网络前向传播和反向传播过程？
- 一个模型的 bias 和 variance 的具体定义是什么？bias 和 variance 哪个比较重要，为什么

是 trade-off? bias 开始的时候很大没问题，为什么开始的时候 variance 也很小？

- 训练网络时有没有出现梯度爆炸？解释一下梯度弥散和梯度爆炸，如何解决。
- 梯度消失怎么解决？为什么会梯度消失？
- 神经网络初始化能等于 0 么，为什么？
- Mini-batch 的作用？

2.2.2 公式推导

- 画了一个神经网络，给了一个输出，让求输出对一个 w 权值的导数，求完导之后问这个输出和这个 w 有没有关系？其他输出和这个 w 有关系吗？
- 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

- RNN 为什么梯度爆炸？
- LSTM 的改进点有哪些？
- 叙述一下 LSTM 三个门，为什么用 LSTM，LSTM 可以用什么代替？
- sigmoid 函数对每个门的作用？

2.3.2 手绘网络原理

- 写一下 LSTM 的公式(给他画了流程图)

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

无

2.4.2 模型评价

- 评估指标、在 AUC 值大时，Log 损失也大。请问可能的原因？

- 问采用什么评价标准，F1 和 P、R 值相比有和不同？

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

- 问 LDA 的原理，应用方面？我就说它是让数据经过映射之后使类间的距离最大化，类内的距离最小，可以应用在分类跟降维。

② 特征选择

- 项目中特征工程怎么做的，从数据流到模型建立？
- 特征工程有什么常用的方法？

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）

- Xgboost 和随机森林各自的特点，差别？分析他们的不同？
- bagging 和 boosting 的区别和异同？

A.基于 bagging：随机森林

- RF 和 GBDT 介绍，RF 在属性的随机采样是有放回的还是不放回的？

B.基于 boosting：Adaboost、GBDT、XGBoost

- RF/XGboost/GBDT 区别，树模型如何处理缺失值和连续值，模型融合 stacking？
- 讲一下 GBDT 和 xgboost 的区别？
- GBDT+LR

② 逻辑回归 LR

- LR 的极大似然函数是凸函数吗?
- 逻辑回归有什么特点? 如果使用逻辑回归, 你是如何提升模型性能的?
- LR 讲一下?

③ SVM (支持向量机)

- SVM 与 LR 的区别?
- 为什么 SVM 做二分类的效果很不错?
- 给我讲一下 svm 的流程, 不要背书, 我看得出来哦。 损失函数是什么, w 的二范式是做什么的, 什么是拉格朗日对偶, 需要满足什么条件, 为什么要求对偶问题? 如果是非对偶的情况呢?

2.5.1.4 无监督学习-聚类方面

- 讲一下 k-means 的原理以及过程?
- K-means 聚类这种方法一定会收敛嘛? 如果不收敛, 你怎么办?

2.5.2 手推算法及代码

- 手撕 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

- 逻辑回归的损失函数是什么?
- 逻辑回归为什么不用 MSE 做损失函数?

2.6.2 网络优化梯度下降方面

- 除了梯度下降还了解什么优化算法? 牛顿法和拟牛顿法, 拟牛顿法的原理以及优势?

2.6.3 正则化方面

- 正则化 L1 和 L2 的原理以及适用场景?

- L1 和 L2 范数各有什么特点以及相应的原因？

2.6.4 过拟合&欠拟合方面

- 如何解决过拟合、欠拟合？

2.6.5 其他方面

- 机器学习和深度学习的区别？

3 vivo 面经涉及项目知识点

第三节
vivo面经
项目知识点
(整理: 江大白)
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

- 画具体的 faster rcnn 网络结构，然后让具体介绍 rpn，fpn 网络？

3.1.2 超分辨

- 知道哪些超分辨率算法？

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 的未登录词怎么处理？
- Bert 的残差网络在哪用到的？
- 问一下你对 bert 的理解，bert 和 transformer 什么关系，bert 的位置编码和 transformer

有什么不同，为什么有这样的不同？

- bert 为什么需要多头，为什么 bert 有 12 层 encoder，如果是 QA 问题，你知道该如何调整 encoder 的层数吗？
- 知道 self-attention 吗，讲一下。如果 bert 中去掉 self-attention 层，还可以拿到词嵌入么，为什么？
- 知道 albert 么？讲一下 albert 的两个改进，对计算量有影响么，为什么？

② Transformer

- Transformer 用的是哪种 attention 机制？
- 画一下 Transformer 结构图

③ Attention

- Self-attention 的 Query, Key, Value 分别是什么。乘积是什么和什么的 Query 和 Key 相乘？
- Self-attention 的乘法计算和加法计算有什么区别？什么时候乘比较好，什么时候加？
- 多头注意力机制的原理是什么？

④ HMM 隐马尔科夫模型

- HMM 具体怎么做的？

⑤ Word2vec

- Word2vec 具体怎么得到词向量？
- 问 Word2vec，为什么没有预训练，word2vec 和 bert 的区别，和 ELMO 的区别？
- 简历中写了 word2vec，问了很多很细的问题

⑥ 其他

- 有没有做过命名实体识别，CRF 了不了解？
- 讲一下 CTC 的原理、CTC 在测试的时候怎么搜索的，两种方式？
- 基于项目，讲一下基于 WFST 的静态解码网络的语音识别流程？除了 GMM-HMM，你了解深度学习在语音识别中的应用吗？

- 给了一个场景，求短文本语义相似度
- 文本的数值化方式，例如：TF-IDF、BOW、One-Hot、分布式的表示方式（word2vec、Glove）等？
- 问 kaggle 情感字段抽取具体是怎么做的，用什么模型，如果句子中出现多个情感字段，怎么解决？如果使用 lstm+crf，当成序列标注来做可以吗，有什么优劣势？

3.3 强化学习

无

3.4 机器学习方面

3.4.1 推荐系统

- 你了解过最新的推荐算法有哪些？推荐系统的 LFM 是什么？解释一下。
- 介绍了一个 CTR 项目，针对项目提问题，特征选择，特征工程，随机森林是怎样重要筛选特征的？
- deepFM 和 wide&deep 有什么区别？wide&deep 是什么样的结构？（补充：你在使用 deepFM 的时候是如何处理欠拟合和过拟合问题的？）
- FM（factorization machine）模型的公式写一下，模型解决了什么问题？
- 手写 deepfm 算法，说明网络结构设计的思路，以及为什么这么做的好处？

4 数据结构与算法分析相关知识

第四节
vivo面经
数据结构与算法分析
(整理: 江大白)
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

- 有序数组的二分查找（重复数字返回第一个），二叉树相加，是否子结构？
- 给你一个数组，让你把零全部挪到数组末尾，并且不改变元素的相对位置？
- 两个数组公共元素最大值？

4.1.1.2 链表

- 手写反转链表
- 如何判断单链表中是否有环？
- 什么是链表，循环链表和链表的区别，循环链表在生活中的例子？
- 输入一个 sorted 链表，删除其中重复的值，return 更新之后的链表？

4.1.1.3 字符串

- 在字符串 abc 求其所有顺序排列 null,a,b,c,ab,ac,bc,abc?
- 给你一个字符串，让你写出所有元素的组合？
- 求一个字符串的所有子序列？(字符串如果有重复字符，只取一次子序列)

4.1.2 树

- 二叉树知道吗，二叉树搜索时间复杂度？和什么有关？
- 求树的深度
- 怎么获得一棵完全二叉树的最后一个结点，时间复杂度，空间复杂度？

4.1.3 排序

- 手撕归并排序
- topk，我给他写了快排的，推了一下快排和堆排的时间复杂度？

- 给了一个栈，里面元素乱序，给了一个相同大小的空栈，和一个变量的位置，让写一个程序对这个进行排序？

4.2 算法思想实战及智力题

- 找出小于 n 的所有质数：筛选法，时间复杂度 $O(n \log n)$ ？
- 最长递增序列

4.3 其他方面

4.3.1 计算几何

- 傅里叶变换推导？

4.3.2 概率分析

- 6 位数字的 8 位数数码管显示的数字，倒过来看和以前相同的概率是多少？
- 8 支球队循环赛，前四名晋级。求晋级可能性？
- 有 100 亿个 email 账号，来一个新账号，先查找是否存在，如果不存在就插入，如何做使得效率最高？允许一定的查找误差率。

4.3.3 其他

- 输入 n ，输出小于 n 的所有质数？

4.4 Leetcode&剑指 offer 原题

- Leetcode 56
- Leetcode 189

5 编程高频问题：Python&C/C++方面

第五节
vivo面经
编程高频问题
(整理: 江大白)
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 网络框架方面

- tensorflow 数据怎么加载到网络里面？
- tensorflow session 是怎么实现的？
- tensorflow 如何数据读取？

5.1.2 基础知识

- dic 和 list 的区别，为什么遍历 dic 时间复杂度是 $O(1)$

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 内存相关

- 变量的内存存储在什么地方？

5.2.1.2 讲解原理

- C++构造函数的作用，什么时候用？
- 析构函数的作用，析构函数要什么形式，为什么要虚函数?虚函数的作用？
- 面向对象的特点，多态是什么意思，怎么实现？
- 静态变量，静态局部变量？
- C++基础，static 和数据段？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
vivo面经
操作系统高频问题
(整理: 江大白)
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

6.2.1 TCP 协议相关

- TCP 和 UDP 的区别？

6.2.2 线程和进程相关

- 线程有了解吗？
- 线程和进程的区别，如何实现多线程？

7 技术&产品&开放性问题

7.1 技术方面

- 只有正样本，没有负样本？怎么解决
- 场景题：有很多数据，它们是以向量形式存储的，怎么把它们自动的进行分类？(聚类)

7.2 产品方面

- 项目中的用户画像你们是怎么做的，怎么分析的？

11|招银网络算法岗武功秘籍

1 招银网络面经汇总资料

第一节
招银网络面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 招银网络面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 招银网络面经-49 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【机器学习算法实习岗】

(2) 全职岗位类

【数据分析工程师】、【大数据研发工程师】、【NLP 算法工程师 (客服机器人、金融风控相关)】

1.3 面试流程时间安排

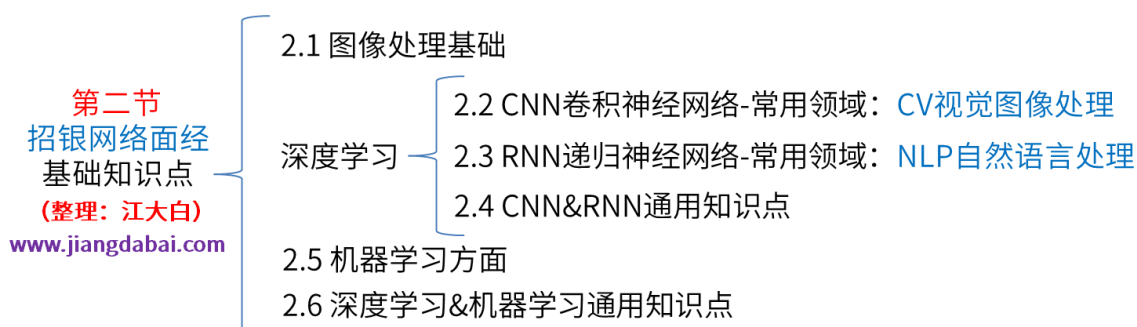
招银网络面试流程 整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	项目技术为主
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	专注项目和领域方向
第三面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

1.4 招银网络面试心得汇总

- ★ 总体来说，感觉题目都比较基础，不过因为是开发的小哥面试的我，所以机器学习和深度学习的问题没怎么问。
- ★ 针对项目问，问得比较宽泛，算法题有的人会考

2 招银网络面经涉及基础知识点



2.1 图像处理基础

边缘检测的原理说一下？

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 卷积方面

- 讲讲 CNN 的各个层，卷积层是怎么实现的？
- 卷积核大小如何选择？

答：大卷积核和小卷积核：大卷积核可分解为多个小卷积核，减少计算量，同时 inception 将 $n \times n$ 小卷积核又分解为 $1 \times n$ 和 $n \times 1$ 的卷积，最后 concat。面试官第一次追问：大卷积核为什么不用于高层。这里没答好。只说了高层大卷积核可能没用，因为没有那么大的特征图来提取，多数是 padding 区域。面试官：还有嘛？第二次追问，卷积核能否偶数。答：不能，会使特征图产生偏移。

- CNN 卷积层的作用？
- CNN 的 padding 有什么用？

2.2.1.2 池化方面

- 解释一下池化层？池化层作用？
- max pool 反向传播？

2.2.1.3 网络结构方面

- 对残差网络的理解，为什么有效？

2.2.1.4 其他方面

- 解释一下 BN？
- 梯度消失、梯度爆炸原因和解决方法？

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- LSTM 的结构讲一下？为啥比 RNN 要快？
- LSTM 长短期记忆如何实现？

2.3.2 手绘网络原理

- 手写 LSTM 的公式（手画 LSTM 图）

2.4 深度学习：CNN&RNN 通用的问题

无

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

- 项目中的数据处理这部分用的啥工具（楼主答 numpy 和 pandas）
- numpy 和 pandas 的实际应用场景以及哪个速度更快一些？

2.5.1.2 特征工程

无

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）

- Bagging 和 boosting 的区别？
- 随机森林和 gbdt 谁主要降低偏差和方差？
- 解释下 RF 和 GDBT？RF 的属性采样时有放回还是不放回？

A.基于 bagging：随机森林

- 讲讲随机森林的原理？介绍一下随机森林？
- 随机森林的重要性排序原理？

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

- xgboost 和 gbdt 区别？
- boosting 了解吗？讲讲流程
- Lightgbm 对缺失值的处理方法？
- Xgboost 为何能并行化？

② 线性回归

- 线性回归为什么用均方误差作为损失函数？

③ 逻辑回归 LR

- 讲一下 LR 的原理？
- LR 和 SVM 的区别？
- LR，从二项分步，指数簇开始讲了下原理，以及为什么使用交叉熵。
- LR 的缺点？

④ SVM（支持向量机）

- 讲讲 SVM，通过什么解原始问题？KKT？
- 核函数，你怎么理解？有哪些核函数？
- SVM 做多分类怎么做？
- SVM 的原理和目标 、SVM 如何防止过拟合？
- SVM 有什么损失函数？
- 问 SVM 与贝叶斯的优缺点？

⑤ 朴素贝叶斯 (Naive Bayes)

- 贝叶斯分类器的原理？

⑥ 决策树 (DT)

- 决策树原理，常用哪几种？
- 决策树相关问题，影响决策树效果的有哪些，开放性问题？
- Adaboost 和决策树的区别？
- 决策树的构建，决策树怎么防止过拟合(前向剪枝，后向剪枝) ？
- 说一下决策树的 c4.5？

2.5.1.4 无监督学习-聚类方面

- 聚类算法有哪些，密度聚类，怎么评价聚类的效果？
- KMean 了解吗？讲一下，KMean 的缺点，以及改进？
- KMeans 和 dbscan 原理，区别，如何调参？

2.5.2 手推算法及代码

- LR 的推导以及应用细节？
- 写出 LR 的损失函数？

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

- 常用损失函数/激活函数？

2.6.2 激活函数方面

- 常用的激活函数说一下？
- Sigmoid 优缺点，对比 Relu 优缺点，怎么改进？
- 激活函数的作用，加在卷积层还是池化层？

2.6.3 网络优化梯度下降方面

- 讲一下梯度下降具体是怎么做的(梯度下降的过程)？

- 问梯度下降算法与牛顿算法的优缺点？

2.6.4 正则化方面

- 正则化有哪些？
- 解释 L1、L2？L1 和 L2 正则化的区别？
- L1 和 L2 为什么一个稀疏一个平滑？

2.6.5 过拟合&欠拟合方面

- 什么样的情况容易过拟合？
- 除了 dropout 和正则化还有什么方法可以处理过拟合（降低网络复杂度）？
- 过拟合的解决方案？1.数据层面：增加数据和数据增强 2.模型层面：换用小的模型，Dropout, L1L2 正则化，以及模型剪枝 3.模型集成以多任务学习

3 招银网络面经涉及项目知识点

第三节
招银网络面经
项目知识点
(整理: 江大白)
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

3.1.1.1 讲解原理

- yolo3 与 yolo4 的区别？
- 小目标检测有哪些改进方式？

3.1.1.2 手写代码

- NMS 算法实现

- 为什么会需要用到 NMS?

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 Word2vec

- Word2vec 如何实现，实现方法有什么区别？

② 其他

- 词袋模型是什么？
- 讲解算法模型 (seq2seq) 以及 TextRank 算法？
- tfidf 公式？
- 基于匹配的问答系统的关键技术是什么（文本相似度匹配）
- 如何判断两个词相似？

3.3 强化学习

无

3.4 机器学习方面

3.4.1 推荐系统

- 基于物品的协同过滤讲一下？
- 介绍推荐系统的召回和排序系统，召回系统的输出是什么？

4 数据结构与算法分析相关知识点

第四节
招银网络面经
数据结构与算法分析
(整理：江大白)
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

- 数组相邻元素之差绝对值都为 1，找给定元素 t 在数组中位置？
- 找出数组中超过一半的那个数？
- 给你两个有序数组，合并排序一下？
- 给定一个数组，找出能构成差为 1 的最长的等差数列的元素。除了先排序后暴力之外，有没有其他方法？
- 给一个数组，已知第一个输出的数，输出栈可能的情况？
- 数组和链表的区别？

4.1.1.2 链表

- 单向链表的倒序输出（vector 再 reverse，或者存在栈里一个个 push 即可）
- 链表反转、链表判断有没有环
- 数组链表的区别？

4.1.1.3 字符串

- 阿拉伯数字汉字转数字，开始以为是 int，被提醒了一下用 string
- 找出一个字符串中第 M 个出现 N 次的字符？
- 字符串查重

4.1.2 树

4.1.2.1 二叉树

- 完全二叉树和满二叉树的区别？
- 二叉树的种类，介绍平衡二叉树

- 搜索二叉树的特性？（左子树比根节点小，右子树比根节点大）
- B 树和 B+树了解吗？讲一下？
- 红黑树讲一下？
- 怎么获取二叉树的深度？我先反问二叉树有没有什么约束，比如完全二叉树之类的，他说没有。之后我就把暴力解法，平衡二叉树，完全二叉树所有的都说了一遍。
- 说一下树的遍历
- 树的深度遍历、广度遍历
- 中序遍历二叉树
- 怎么查到第 k 个大的数（中序遍历二叉搜索树）

4.1.2.2 堆

- 堆和栈的介绍以及它们之间有什么区别？

4.1.3 排序

- 常用排序算法、哪几种排序算法时间复杂度比较低、讲堆排序原理？
- 介绍一下排序算法，选择一种(除了冒泡)进行具体介绍，它是怎么操作的？
- 针对你具体讲的那种排序算法，追问你时间复杂度分析以及空间复杂度分析？
- 几种排序复杂度稳定性
- 讲一下归并排序
- 海量数据排序，我写的堆排序
- 写快排和堆排，然后比较

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 青蛙跳台阶

4.2.2 智力题

- 8 个砝码，一个超重，用天平最少几次称出来？（2 次）

4.3 其他方面

4.3.1 矩阵运算

- 手撕螺旋矩阵
- 给定矩形内有若干点，求距离最小的两个点的距离，不能用穷举法？
- 给一个二维矩阵和一个值，怎么判断这个值在矩阵中，矩阵很大，写出你能想到的最优方法？

4.3.2 其他

- 贪心算法的原理

4.4 Leetcode&剑指 offer 原题

- Leetcode 原题：怎样实现一个 LRU，说思路以及用到的集合框架
- Leetcode 原题：求一个字符串里面最长的回文串
- 剑指 offer 第 33 题：数组排成最大的数字

5 编程高频问题：Python&C/C++方面

第五节
招银网络面经
编程高频问题
(整理: 江大白)
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 网络框架方面

- 描述一下 TensorFlow 计算图，工作原理、用的什么数据结构？

- Tensorflow 数据流图?

5.1.2 基础知识

5.1.2.1 内存相关

- Python 分配内存方法

5.1.2.2 区别比较

- tuple 与 list 的区别?
- Python 元组和列表的区别?
- not exists 和 not in 区别?
- 生成器和迭代器区别?

5.1.2.3 讲解原理

- 知道 Python 的匿名函数吗?

5.1.2.4 讲解应用

- 说一下 Python 的 yield: 生成器, 惰性机制
- 讲一下 Python 装饰器
- Python 的几种运算符介绍?
- Python new & init 方法
- 常用的 Python 包

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 线程相关

- 多线程的切换是怎么实现的?

5.2.1.2 内存相关

- 讲述一下 C++内存泄漏和内存溢出?
- 内存对齐讲一下?

5.2.1.3 区别比较

- 引用和指针的区别?(会深入问为什么)
- 讲述 C++中重载和重写及区别?

5.2.1.4 讲解原理

- 介绍一下构造函数，析构函数，函数重载?
- 多态怎么实现?
- 介绍一下 C++编译过程
- 索引的优缺点
- 面向对象了解吗? 讲讲主要特性?

6 操作系统高频问题：数据库&线程&常用命令等

第六节
招银网络面经
操作系统高频问题
(整理: 江大白)
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

6.1.1 基础问题

- 数据库索引的基本原理

6.1.2 手写代码

- 一个三个字段的数据表, id,name,age, 写一个 where 查询, 写一个按照年龄分组求每组有多

少人。直接用 case when 分组再 group by 后 count?

- sql 模糊匹配
- sql 里面用过的 join 语句

6.2 操作系统方面

6.2.1 线程和进程相关

- 进程与线程的区别?
- 深拷贝和浅拷贝?

6.2.2 常用命令

- Linux 复制命令是什么? 查询帮助用什么命令?

7 技术&产品&开放性问题

7.1 技术方面

- 深挖任务型对话系统的细节?
- 多轮对话的场景保存和跳转怎么做? 可以通过什么数据结构来实现?
- 怎么用知识图谱做 qa?
- 遇到脏数据怎么处理?
- 给定一段文本, 如何识别出其中的省市区等地址并正确拼接
- 场景题: 类似于反欺诈识别---给定用户画像, 如住址 (文本) --关键、年龄、职业、消费金额 (num-feature)、其他信息 (文本)
- 如何精确识别欺诈营销?(对文本处理不是特别了解, 大致说了 kmeans 的方案, 有待改进)
- 大数据处理中的数据不平衡和数据倾斜讲一下?
- 说一下信用分预测用的什么模型, 说说模型的原理? (说用了 lightgbm, 讲了下原理)
- 场景题: 交易检测场景, 拿到所有用户的交易数据, 我们知道一些数据是正常的操作数据,

一些是非正常操作的数据，会引发投诉，问我们怎么检测非正常操作数据？（面试官想问的是机器学习的一般步骤）

- 深度学习和机器学习，讲讲你的理解？深度学习好在哪？缺点在哪？

7.2 产品方面

- 银行怎么预测用户的流失率，你从哪些角度去考虑？我说的是用户持有卡的收入支出和交易频率与以往记录相比是否有下降趋势？

12|360 公司算法岗武功秘籍

1 360 公司面经汇总资料

第一节 360公司面经 汇总资料 (整理: 江大白) www.jiangdabai.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
	1.4 360公司面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 360 公司面经-46 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 全职岗位类

【计算机视觉工程师】、【搜索广告算法工程师】、【机器学习算法】、【语音识别研发】、【搜索部门的机器学习工程师 (偏 nlp 方向)】、【360 效率平台开发工程师 (也问算法题)】

1.3 面试流程时间安排

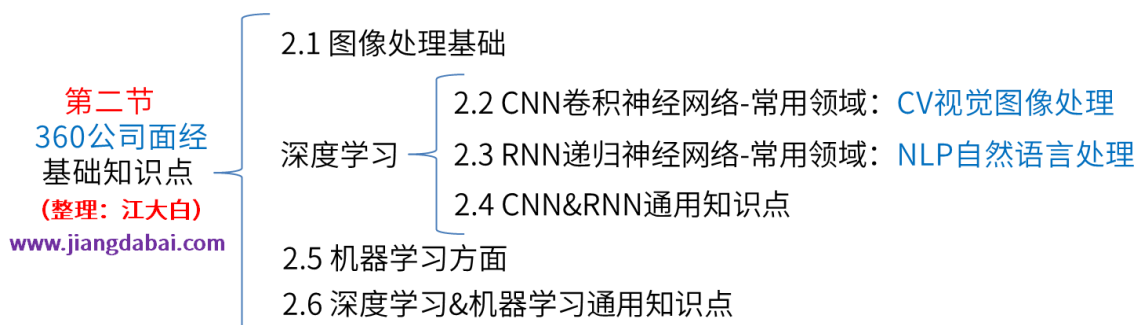
360公司面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第三面	HR面	基础人力问题	/

PS: 以上流程为大白总结归纳所得，以供参考。

1.4 360 公司面试心得汇总

- ★ 对着我的机器学习的项目详细问，这里必须得说，不会的东西千万不要往简历上写。
- ★ 技术面面的比较基础， 比较在意了解算法的广度
- ★ 项目中用了什么方法最好能知其所以然，才能处乱不惊，没准还能想出新的 idea
- ★ 面试算法的同学还是要自己吃透自己的简历，大多数的面试官问题都是出自你的简历
- ★ 受益匪浅，面试官跟我说做这一行最重要的还是从数据出发，分析数据，从数据反应出的现象去寻找问题，再思考原因，然后去做出自己的假设，假设有可能对，有可能错，但这也正是做算法的魅力之处。有经验的算法工程师的假设都是建立在对数据和对模型底层特别了解的情况下。
- ★ 模型不仅仅只是工具，做模型的时候不要一蹴而就，要一点点的加东西，你要去弄懂你为什么加，加了能带来什么改进等等。
- ★ 360 面试时，参照下面面试官提问的思路，一般是从你讲到的点里挑一个自己感兴趣或很了解的点进行下一步提问，所以讲话要特别注意，自己不熟的东西尽量不要穿插在里面。

2 360 公司面经涉及基础知识点



2.1 图像处理基础

- 讲一下卡尔曼滤波算法的公式，影响的参数，怎么调参？

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 卷积方面

- 上采样的方式
- 反卷积和空洞卷积讲一下
- CNN 中怎么做的卷积？
- 卷积的优势？卷积之后接什么层？
- 尝试过 BN 层和卷积层相结合吗？

2.2.1.2 池化方面

- 为什么做 pooling？都有哪些 pooling？

2.2.1.3 网络结构方面

- Resnet 用过吗？
- 介绍 Alexnet 的网络结构？

- VGG 网络相比于 AlexNet 改进的地方在哪里，为什么？

2.2.1.4 其他方面

- BN 原理是什么，实验的收敛速度增大了多少？（提升了 3 倍，我的）
- 神经网络了解过吗，讲一下权值共享？
- 梯度消失/爆炸的原因及解决方法？

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- 介绍 RNN、LSTM、GRU 并比较
 - LSTM 与 RNN 的区别？
 - LSTM 门到底那个门更新细胞状态？
 - LSTM 各个门的作用？我口头讲了一下 LSTM 的计算公式，然后又问为什么 sigmoid&tanh？
- 然后衍生到 LSTM 为了解决 RNN 的什么问题？GRU 是否了解？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 模型评价

- AUC 计算公式？
- 项目为什么选用 AUC 和 F-score，为什么不选用 Recall？
- Precision/Recall/F-score/ROC/AUC 的原理讲一下？
- 多分类的评价指标
- 二分类的评价指标

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

- PCA 理论的过程

② 特征选择

- 机器学习 one-hot 编码了解吗？

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）

- 介绍一下有哪些集成算法？
- XGB 和随机森林的区别？
- XGB 和随机森林哪个的基模型适合更深的树？
- 讲讲 GBDT，XGB 的区别？
- XGB 的损失函数
- XGB 与 LR 相比优缺点在哪里
- 除了 XGB 和 LR 还了解哪些机器学习模型？
- gbdT 原理，怎么实现的？
- Xgboost 的原理介绍以及如何并行化实现？

② 逻辑回归 LR

- 说说对 LR 的理解？
- 既然 LR 解决的是分类问题，它为什么叫逻辑回归？

- LR 和 SVM 的差别?
- LR 和最大熵模型的相似点与不同点?
- LR 的公式, 使用什么方法更新, 推导求梯度的公式
- LR 的损失函数, 为什么要用交叉熵?
- LR 如何做正则化?
- 逻辑回归可不可以加核函数?
- 归一化对 LR 和决策树的影响?

③ SVM (支持向量机)

- SVM 原理讲一下?
- 项目上为什么选 SVM, 用的什么核函数, 了解其他的核函数么? 和 LR 模型区别?
- SVM 的 Loss 是什么, 核函数的作用? SVM 核函数变换有什么好处?

2.5.1.4 无监督学习-聚类方面

- 用过什么聚类算法, 什么时候需要用聚类?
- kmeans 原理? 如何确定 k 值?
- kmeans 有哪些优化方法?
- 讲一下 k-means 算法, 怎么做, 遇到什么问题, 怎么解决的?

2.5.2 手推算法及代码

- LR 推导
- 手写 LR 的目标函数
- 写一下高维高斯分布密度函数?

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

- 写一下交叉熵的公式

- 逻辑回归什么时候用 softmax，什么时候用交叉熵，为什么？

2.6.2 激活函数方面

- Relu 和 Sigmoid 函数之间的差别？（1.不会梯度弥散；2.稀疏参数；3.计算简单）

2.6.3 网络优化梯度下降方面

- 比较随机梯度下降、批量梯度下降、小批量梯度下降？

2.6.4 正则化方面

- 说一下正则化的作用，以 3 种角度介绍了 L1 和 L2 正则化的区别？
- L1 正则化、L2 正则化讲一下

2.6.5 过拟合&欠拟合方面

- 过拟合的原因，如何解决过拟合？

2.6.6 其他方面

- 机器学习和深度学习的区别是什么？（数学完成，经验驱动；数据量少，数据量大）
- 什么样的任务适合用深度学习，什么样的问题不适合？

3 360 公司面经涉及项目知识点

第三节
360公司面经
项目知识点
(整理: 江大白)
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

- 小目标检测方法？

- 是否对新的检测方法有了解？（答了几个 anchor free 的方法）

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

3.2.1.1 讲解原理

① Transformer

- transform 的 mask 到底有什么作用？
- 讲下 transformer 中 self-attention

② Attention

- 讲下 attention 在 seq2seq 中的用法

③ HMM 隐马尔科夫模型

- HMM 了解过吗？有什么样的应用场景？
- 讲语音识别整体流程，然后着重讲声学模型，GMM-HMM，DNN 模型怎么用的。再讲解码过程，怎么计算概率等？

④ Word2vec

- 讲一下 Word2vec 原理？怎么优化？
- 如何用 word2vec 的方式构造 sentence2vec？
- 介绍 word2vec 原理，如何设计，如何训练，越详细越好；用了什么加速方法，word2vec 算法本身有哪些优化，加速方法，最后的一层 softmax 怎么加速？
- Word2vec、elmo、gpt 的区别与联系？

⑤ 其他

- 分词的原理？

3.3 强化学习

3.3.1 讲解原理

- 项目中的 wgan, wgan-gp 怎么实现的?

3.3.2 损失函数

- GAN 网络的 Loss 函数说一下?

3.4 机器学习方面

3.4.1 推荐系统

- 协同过滤算法的原理和应用场景?

4 数据结构与算法分析相关知识点

第四节
360公司面经
数据结构与算法分析
(整理: 江大白)
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

- 快排数组的引用传递和值传递 (是否会改变原始数组?)

4.1.1.2 链表

- 写一个反转单链表的代码, 不要递归?
- 两个有序链表的合并?

- 链表相乘

4.1.1.3 字符串

- 一个基于字符串匹配的中文分词，正向长优先匹配？

4.1.2 树

4.1.2.1 二叉树

- 哈夫曼树与负采样？
- 实现哈夫曼树手撕代码？
- 二分查找
- 任意节点开始的二叉树的路径？
- 二叉树的前序遍历

4.1.2.2 堆

- 堆和栈的区别（内存分配）？

4.1.3 排序

- 有哪些排序算法，复杂度分别是多少？
- TopK（快排和小顶堆分别实现，分析时间和空间复杂度）
- 手写快排
- 求出十万个数据中的前 k 大的值，怎么做，可不可以用堆实现？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 跳台阶的问题，一下一个，或者一下两个，一共 n 阶，问多少种跳法，用递归很简单？
- 给一个环形数组长度为 N，数组每个位置都有一种颜色，总共 M 种颜色，求包含全部颜色的最短长度？

4.2.2 智力题

- 横着放六个硬币，以第一个硬币开头竖着发放四个硬币，所以目前总共十个硬币，横着看六个，竖着看五个，问怎么移动一个硬币可以使得横着看六个竖着看也是六个？
- AB 两个人，背后各有一个数字，两个数字相邻，A 能看见 B 的，B 能看见 A 的，然后 A 和 B 的对话是，A 说不知道，B 说我也不知道，A 说我知道了，B 说我也知道了，问 AB 两个人背后的数字是多少？
- 一个四位数乘以 4，变成它原来反着的数，比如 1234 变成 4321，求这个数是啥？
- 给你一个整数数组，找到从 1 开始缺的最小的数。比如说就是-2, -1,0,3,2,1,5，那从 1 开始缺的最小的数就是 4。

4.3 其他方面

4.3.1 数论

- 实现 shuffle，1000 个数随机选 10 个？

4.3.2 计算几何

- 在平面上所有点中，找一条直线是经过的点最多？

4.3.3 矩阵运算

- 二维矩阵最短路径
- 给定一个二维矩阵，二维矩阵非 0 即 1，随机生成，求能框住所有 1 的最小矩形面积？

5 编程高频问题：Python&C/C++方面

第五节
360公司面经
编程高频问题
(整理: 江大白)
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 网络框架方面

- 项目中用到 Caffe, 你要重写类的话, 要怎么加?引申出来问了 caffe 的类的工厂模式懂么? 为什么 caffe 写了类放在相应目录下就可以? 类的注册?

5.1.2 基础知识

- Python 类的私有变量怎么设?

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 区别比较

- C 和 C++的区别。讲一下多态性?

5.2.1.2 讲解原理

- 关于动态内存, 除了 new、delete 用什么实现动态内存?
- C++中的 STL? C++ const?
- C++中的构造函数哪几种? 对一个 class A{} 写一下构造函数
- 介绍一下 C++的三大特性?

6 操作系统高频问题: 数据库&线程&常用命令等

第六节
360公司面经
操作系统高频问题
(整理: 江大白)
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

6.2.1 线程和进程相关

- 线程和进程的区别，各有哪些资源？
- 了解 linux 内核吗？

7 技术&产品&开放性问题

无

13|海康威视算法岗武功秘籍

1 海康威视面经汇总资料

第一节 海康威视面经 汇总资料 (整理: 江大白) www.jiangdabai.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 海康威视面经整理心得
--	--

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 海康威视面经-64 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 全职岗位类

【大数据算法工程师】、【图像算法工程师】、【数据挖掘工程师】、【自动驾驶算法工程师】、【AI 算法工程师】、【视频编码算法工程师】、【杭州研究院大数据算法工程师】、【智能传感算法工程师】、【AI 算法加速工程师】

1.3 面试流程时间安排

海康威视面试流程 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	简历项目&基础知识
第二面	技术Leader面	自我介绍+项目经验+技术问答	深挖项目，也会问 更宏观、开放一些的问题
第三面	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：

- 两面+HR面+后面可能会加一面，比如大 boss 或者技术大佬面，可能会给 SP 级别
加面全程扣项目细节，面试官如果是数学大佬，一直深挖项目里数学公式的推理证明。其中按照经验值设定的常数，会问你经验值是怎么推导出来的，类似这种，问的很细。
- 有的人在第一面技术面之前，会加电话面，可能是简历初步筛选

1.4 海康威视面试心得汇总

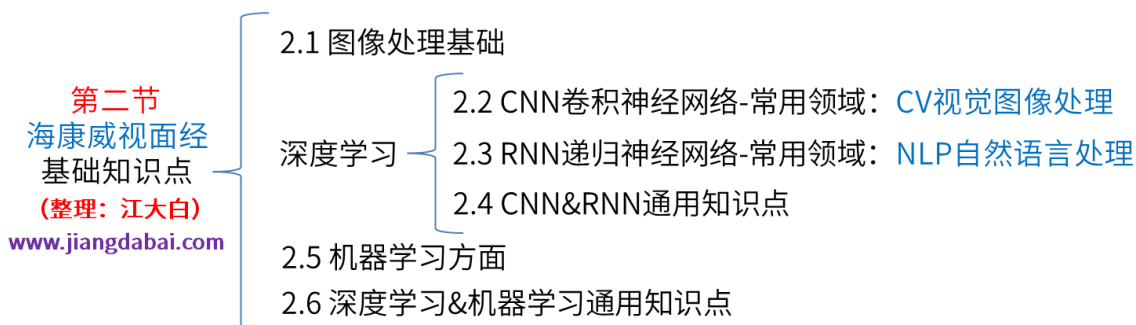
★ 感觉海康不是那么看重相关经历，主要考察你的逻辑思维能力和解决实际问题的能力，应该是看重你的潜力吧。

★ 一面就是一定要对自己的项目非常非常熟悉，遇到了哪些困难，怎么解决问题的，有哪些创新等等

★ 二面面试官应该是个 leader，更关注宏观上的东西，如果在学习的时候只关注自己领域的具体细节，对其他相关领域不关注，可能容易问到知识盲区。讲话逻辑清晰，而且没有一句废话，从项目开始，一点点深挖，比如项目中用到这个方法，是谁提出来的，如果让你用别的方法你会怎么做，如果没有这些数据你会怎么搞，如果这个方法不行了你会怎么处理，都是很抽象的考察。但是细节的地方也会追问，讲完了还会帮你一起用一两句话总结。

- ★ 感觉 HR 面，不像是纯人力资源的面试官，他们好像都懂些技术，或者是对接技术岗的 HR
- ★ 海康面试过程中，中间的 battle 过程，主要还是看你如何把话语主动权尽可能多拉到自己这边，记得扬长避短，面试官双商都很高，一定不要胆怯被问到语塞，要直面问题并迅速抓取他感兴趣的点，尽可能的把自己的优点在不同问题中展现。

2 海康威视面经涉及基础知识点



2.1 图像处理基础

- 怎么理解图像的频率？
- 介绍一下双边滤波？
- 边缘检测算法，膨胀腐蚀的原理介绍一下？
- HOG 特征，如何进行计算？
- 立体匹配方面，给外行讲一下角点检测和立体匹配？

答：角点检测：角点是具有一定特征的点，具体包括在角点处一阶导数最大，二阶导数为 0，其指示了物体边沿变化的方向。根据这些特征，我们可以对图像中的物体进行定位。比如，构造已知具有一定距离的角点物体，通过拍摄获取物体的图像信息，对图像中角点信息的提取，搭建起图像和现时的比例映射关系。其比较经典的应用在于，标定。

立体匹配：对于两张或者多张图片，若图像中存在同一个物体的像点，则找到同一个像点在不同图像中的位置差异，即为立体匹配的过程。

- 图像缩放算法有哪些？
- 图像插值算法之双三次插值的原理讲一下？

- 空间域处理-空间滤波-平滑空间滤波器了解吗？
- 图像高斯模糊运算直观解释一下？
- 把一个图像旋转一定角度，要怎么做（关键是插值这个点）

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 卷积方面

- 可变形卷积在实际部署上有什么问题？
- 卷积提取的信息是什么？

2.2.1.2 网络结构方面

- Resnet 相比于之前的卷积神经网络模型中，最大的改进点是什么？
- Resnet 之后有哪些新的基础网络，介绍一下？（resnext, SENet, DenseNet 等）
- InceptionV2、V3 的区别了解吗？

2.2.1.3 其他方面

- 感受野是如何计算的？
- BN 的原理和作用？
- Batchnormalization 在训练和测试的时候有什么不同（项目介绍提到过）？

2.2.2 激活函数类

- 讲一下 softmax？
- 详解 softmax 函数以及相关求导过程？

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- LSTM 网络参数数量计算公式的推导？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 问了一些对于 CNN 现在强势抢占 RNN 应用场景（机器翻译，语音合成等）背后的 insight？
- 正负样本不均衡时的解决方案？

2.4.2 模型评价

- 模型如何进行评价？分类：P-R 图，F1 值
- 样本不均衡，只对大样本有兴趣，怎么评价？
- F1 值的原理讲一下？
- 准确性、精确率、召回率的指标是如何判断的？
- 准确性、精确率、召回率是不是随着置信度的大小而改变，你们是怎么设置这个置信度的，有没有试过别的阈值？

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

- 数据清洗：异常值怎么判断？箱型图：上界怎么定义？极端异常值怎么判断？

2.5.1.2 特征工程

① 特征降维

- PCA 理论过程？

- 简单介绍一下 PCA?
- PCA+SAE 和 SAE+PCA 有什么不同，分别做过对比吗?

② 特征选择

无

2.5.1.3 有监督学习-分类和回归方面

① 分类回归树

- 随机森林的随机是什么，是样本随机采样还是特征随机采样?
- GBDT 和 XGBoost 的区别，XGBoost 都调过哪些参数，这些参数有什么作用，效果怎么样?
- XGBoost 和 lightGBM 区别，各自优势，调参心得?
- GBDT 原理、在处理 onehot 属性时，当 1 或者 0 的数量很多的时候是否会发生问题?既然 GBDT 处理 onehot 属性时存在问题，为什么你们还要这么做?
- gbdt、xgb、lgb 的区别和具体推导?
- xgb 的一些问题
 - (1) 对 gbdt 的改进?
 - (2) 用了什么提高速度的方法，怎么并行化的?
 - (3) 为什么用了泰勒二阶展开而不是三阶四阶?
 - (4) xgb 里面怎么处理缺失值?
 - (5) xgb 直方图算法怎么做的?

② SVM (支持向量机)

- SVM 的原理? 什么样的函数是核函数，如何去选择核函数，线性，高斯核函数。

③ 决策树 (DT)

- 决策树的分割点的选择方法有哪些?

2.5.1.4 无监督学习-聚类方面

- kmeans 算法，em 算法，推导 kmeans 算法收敛的条件?

- 简单介绍一下 Kmeans, K 值选取的方法, 初始中心怎么优化?

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

- 常见的 Loss, 回归的, 分类的?
- 如何通俗的解释交叉熵与相对熵?
- 交叉熵公式推导?

2.6.2 激活函数方面

- 怎么理解 Relu 函数的作用?
- 写了 sigmoid 及其导数, 还有他的值域?

2.6.3 网络优化梯度下降方面

- Adam 和任意一个变种的推导式?

2.6.4 正则化方面

- 有哪些正则化方法?
- 正则化项的功能作用?
- 写一下 L1、L2 的数学表达式

2.6.5 压缩&剪枝&量化&加速

- 优化模型容量的方法, 先排除说简化 batch, 优化模型结构和压缩模型, 使用半精度或者多 gpu 这些外, 还问还有啥技巧?
- 围绕项目, 讲 int8 模型量化和卷积加速的多种方式(fft, winograd, im2col+sgemm)

2.6.6 过拟合&欠拟合方面

- 什么是过拟合? 防止过拟合有哪些操作?
- 如果发生过拟合怎么办? 改善过拟合的办法?

- L2 正则化和 dropout 为什么可以防止过拟合?
- 介绍模型压缩, 模型蒸馏 (分类与目标检测两个方向)

3 海康威视面经涉及项目知识点

第三节
海康威视面经
项目知识点
(整理: 江大白)
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

3.1.1.1 讲解原理

- 目标检测的几个主流深度算法以及优势?
- 给两个 bbox 怎么求 iou? 多边形 IOU 计算为什么不均匀撒点?
- 说一下 faster-rcnn 的整个从输入到输出的框架流程、说一下 rpn 的原理?
- 目标检测中针对小目标的解决措施
- 如何解决类内的检测?
- Two-stage 的检测和 One-stage 的区别?
- RCNN, Fast-RCNN, Faster-RCNN 的发展历史?针对 Faster R-cnn 的问题, 如何解决后面 FC 过多的情况?
- 检测的框角度偏移了 45 度, 这种情况怎么处理?
- 说一下 anchor-free 模型 FCOS 的原理和损失函数?
- FPN 以及 SSD 算法
- 介绍 anchor 与 anchor free 的优缺点

3.1.1.2 损失函数

- Focal loss 具体怎么操作的、说下 focal loss 中两个参数的含义，以及它存在的风险？（风险这块商汤面试的时候也问了，我似乎都没答全，准备去补一补）

3.1.2 图像分割

- 现在最好的分割模型是什么，我答了 deeplab v3+，面试官接着问，deeplab v3+相比于 deeplab v2 的区别在于什么？

答：deeplab v2 也是基于 encoder 和 decoder 架构的，但是后面有连接 CRF 条件随机场，而 deeplab v3+ 没有了条件随机场，deeplab v3+ 的亮点之处在于引入了 ASPP 空洞卷积模块和同步的 BN。

- 分割出来的结果通常会有不连续的情况，怎么处理？

3.1.3 超分辨

- 解释一下什么是超分辨率？（跟个人研究课题有关）

3.2 深度学习：RNN 递归神经网络方面

无

3.3 强化学习

- 强化学习从 dp 到蒙特卡洛方法、TD 差分法到最后的 DQN 各个阶段的原理和 trick 都介绍一遍？（期间会打断问一些比如 off policy 和 on policy 区别之类的问题）
- 强化学习未来的发展路线，和图像、语音这些如何结合？
- PPO 相对于 TRPO 的改进？
- 什么情况下使用重要性采样？
- 直观上来说，AC 方法是 on-policy 还是 off-policy？既然是 on-policy 的方法，A3C 没有使用重要性采样的情况下，直接取之前策略的数据会不会有问题？
- DP 方法和 TD 方法区别在哪里？面试官说这个涉及到本质问题。

- 为什么要采用优势函数？

3.4 机器学习方面

无

4 数据结构与算法分析相关知识

第四节
海康威视面经
数据结构与算法分析
(整理: 江大白)
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

- 寻找乱序数组中第 K 大的数？
- 一组乱序数组，一直给进输入新的数，求整体数组的中位数，时间复杂度？

4.1.1.2 链表

- 两个有序链表排序，k 个有序链表排序，时间复杂度？
- 2 个有序链表的合并到多个有序链表的合并？PS：给出最优解法、并非暴力法

4.1.1.3 字符串

- 字符串翻转 (abcdef -> efabcd)？
- 手撕代码 (123->321)
- 给定一个字符串，找出最长的回文子串（只会判断是否为回文子串的我，只说了暴力解法思路，代码都没写）

4.1.2 树

- 二叉树路径和?
- 二叉搜索树中寻找给定值? (每个节点的 val 不一样) 就是和本节点值比较, 然后决定是否找到/分到左子树还是右子树。

4.1.3 排序

- 冒泡和快排、问了如何优化? 冒泡就是加个计数器, 为 0 就 break 循环, 快排就是随机选取 pivot。
- 快排和堆排的优缺点和应用场景?
- n 个数取 topk?
- 给一个数组排序, 不限排序方法, 写了快排, 问平均时间复杂度, 然后问有没有更快的排序方法?

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 青蛙跳台阶

4.2.2 智力题

- 买 n 个苹果, 6 个一袋, 8 个一袋, 求最少买几袋?
- 一个公司的员工 X 人, 平均年龄 Y 岁, 每年离职率 N, 新员工都是 21 岁, 每年都会补到 X 人, 求 P 年后公司的平均年龄?
- 一根绳子对折 n 次, 从中间剪一刀, 共分成几段?

4.3 其他方面

4.3.1 数论

- 三个整数 a, b, c, 怎么找到中间值?

- 给定一个正整数 n ，问 n 最少可用多少个整数的平方和表示？（四平方和定理，故只需要考虑 n 是否可以用 1、2、3 个整数的平方和表示。当时没有想出判定 n 是否可用 2、3 个整数平方和表示的数学解法，直接使用的 $O(n)$ 的暴力解法。但也可使用更优化的方法）

4.3.2 概率分析

- 从一副 52 张扑克牌中随机抽两种，颜色相等的概率？
- 一根筷子掰两次，组成三角形的概率？
- 一根绳子随意砍两刀，得到三段，组成三角形的概率？

4.3.3 矩阵运算

- 给两个矩形的左上角顶点和两个矩形的变长，判断矩形是否相交？

4.3.4 其他

- 斐波那契数列，写出空间复杂度为 1 的求法？
- 从左上角到右小角最短路径？
- 堆与栈的区别

4.4 Leetcode&剑指 offer 原题

- Leetcode72：字符串编辑距离 有没有比动态规划更好的方法？
- 剑指 Offer15：求一个数二进制表示中 1 的个数，问了负数怎么处理

5 编程高频问题：Python&C/C++方面

第五节
海康威视面经
编程高频问题
(整理: 江大白)
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 网络框架方面

5.1.1.1 Pytorch 相关

- Pytorch 的 model 流程讲一下？

5.1.1.2 Caffe 相关

- 对 caffe 中 blob 的理解？（项目中使用到的）

caffe 中的 blob 跟 tensorflow 中的 tensor 一样是一种存储数据的矩阵，在网络中的数据的传递和计算就是用的 blob，一般对于做图像相关领域来说 blob 是 4 维的。

5.1.2 基础知识

5.1.2.1 区别比较

- python 里面的浅拷贝和深拷贝

5.1.2.2 讲解原理

- Python 类的私有变量怎么设？
- python 里的迭代器了解吗？

5.1.3 手写代码相关

- 现在有一个 list，怎么把它排序？（sorted 函数）

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 内存相关

- python 的深拷贝和浅拷贝

5.2.1.2 区别比较

- 结构体和联合体的区别是什么？
- 引用和指针的区别？

5.2.1.3 讲解原理

- 虚函数、多态的意义？
- C++熟悉吗？解释下 C++的面向对象的思想
- 有几种类型转换的方法？（隐式转换，显式转换）
- C++ 继承？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
海康威视面经
操作系统高频问题
(整理: 江大白)
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

6.2.1 常用命令

- Linux 常用命令熟悉吗？--面试官会先问是否熟悉，如果不熟悉的话，就不会再问。如果熟悉，就会相应地问一些内容。据面试官讲，并不要求全都会，重点看潜力吧。
- Linux 系统下查看 CPU 使用情况？
- 计算文件中所有 jpg 图片的数量的命令？

7 技术&产品&开放性问题

无

14|快手算法岗武功秘籍

1 快手面经汇总资料

第一节
快手面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 快手面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 快手面经-77 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【快手杭州商业化算法实习生】、【算法测试实习生】、【用户增长算法实习生】

(2) 全职岗位类

【计算机视觉工程师】、【机器学习算法工程师】、【广告算法工程师】、【openday 社区科学部算法工程师】、【推荐算法工程师】、【图像增强算法工程师】、【y-tech 部门工程师】、【数据挖掘算法工程师】、【视觉识别算法工程师】、【商务化业务部广告算法工程师】、【语音算法工程师】、【快

手游戏 AI 算法工程师】、【NLP 算法工程师】

1.3 面试流程时间安排

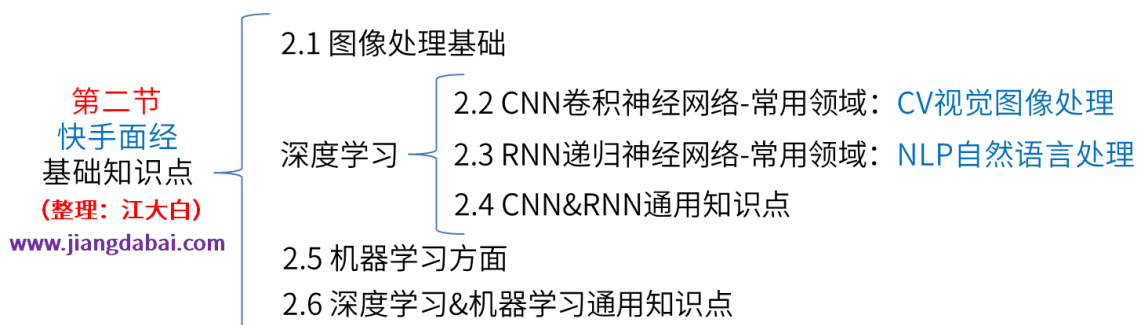
快手面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	有的考官会深挖项目 细节及创新点
第三面	技术Leader面	自我介绍+项目经验+公司发展	/
第四面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

1.4 快手面经面试心得汇总

- ★ 对简历上自己实习或者做过的项目的细节，要知道清楚，会针对项目问得很深
- ★ 二面和面试官交流了很多，深刻认识到数学在算法方面的重要性，“数学能力决定上限，代码能力决定下限”。
- ★ 感觉快手都没咋问基础知识，基本就是项目经历+算法，感觉比头条简单一些。
- ★ 还是得多刷算法题，常见题目必须得秒写，二叉树，链表，二分查找，这是目前这么多面试出现频率最高的题目，可能是因为实习生的要求不高，问的都比较基础，估计校招的时候就会问一些 DP 相关的题目了。
- ★ 有人说的 HR 面的问题，基本都是一模一样，先介绍部门然后问我什么城市选择、职业规划、三个词形容自己等。

2 快手面经涉及基础知识点



2.1 图像处理基础

- 手写直方图拉伸?
- 手写均值滤波?

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 卷积方面

- 什么是 dropout? 让你实现一个 dropout 你怎么实现?
- dropout 原理讲一下? 作用以及训练测试时的不同?

2.2.1.2 池化方面

- 池化层的作用? 池化层反向传播的时候是怎么计算的?

2.2.1.3 网络结构方面

- 介绍常用的网络结构, 如何改进?

2.2.1.4 其他方面

- BN 层解释一下? 有啥用为啥好用?
- BN 的原理、作用以及训练测试时的不同?

- BN 为什么能加快收敛速度?BN 解决了什么问题? 是怎么计算的? 基于什么计算的? 训练和测试的时候有什么不同? 测试时候的均值和方差是怎么来的?

- 为什么输入网络之前数据要做归一化?

- 梯度爆炸的原因及解决方法

- BN 的 γ β 意义

2.2.2 数学计算

- 卷积时间复杂度?

2.2.3 激活函数类

- 介绍 softmax, 对 softmax 的理解

- 证明 softmax 的特征规划问题, 在撕代码的界面推

2.3 深度学习: RNN 递归神经网络方面

2.3.1 讲解相关原理

- LSTM 各类门结构?

- LSTM 的门是什么, 激活函数是什么?

- RNN 的反向传播 + bp 推导

2.4 深度学习: CNN&RNN 通用的问题

2.4.1 基础知识点

- 训练集测试集和验证集怎么划分?

- 训练数据太少怎么办?

- 模型 attention 怎么做, 结果效果?

- overfitting 的解决方法 为什么会产生 overfit 的现象

- 介绍一下 transformer 的结构？encoder 的输入？decoder 的输入？介绍一下 self attention？为什么比 RNN 好？encoder 和 decoder 都能并行？decoder 的 mask？

2.4.2 模型评价

- AUC 啥意思？横纵坐标什么？物理含义什么数据分布改变时，AUC 有什么影响（没啥影响，我居然说我不太清楚？）
- AUC 的定义和计算方法、auc 实际上衡量的是什么能力，roc 曲线是否会出现先增后减的情况？

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

- 介绍一下自己在数据清洗这一块有什么心得方法？

2.5.1.2 特征工程

无

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）

- gbdt 和 rf 的区别？
- 讲一下 rf 和 lightgbm 区别？
- 集成学习的方法有哪些？

A.基于 bagging：随机森林

- 分析一下随机森林可不可以剪枝？
- 随机森林是什么原理？

B.基于 boosting：Adaboost、GDBT、XGBoost

- 介绍一下 LightGBM 与 Xgboost 的区别？Lightgbm 相对于 xgboost 的优化？

- 介绍一下 GRU、RF、GBDT、XGBoost
- 如果判断树模型 lgb 过拟合，怎么调整树的参数？
- XGBoost 的特征重要性是如何得到的？
- GBDT+LR
- 问了 xgboost 缺失值处理方法？
- 为什么 xgboost 要二阶展开？

② K 近邻 (KNN)

- KKT 条件具体是什么？

③ 逻辑回归 LR

- 问了标准化，那些模型需要对数据进行标准化，标准化有什么用，LR 中是否一定需要标准化？

④ SVM (支持向量机)

- 通俗的讲一下 SVM？
- 讲 SVM 原理，SVM 损失函数是什么？
- SVM 为什么能够求解对偶问题，求解对偶问题为什么和原问题一样？为什么要求解对偶问题？svm 的公式是什么？如果线性不可分怎么办？
- 问了支持向量机的基础思路，支持向量机的损失函数，了解支持向量机的核函数吗？核函数需要满足什么样的条件？核函数怎么进行构造？
- SVM 解决线性问题还是非线性问题？

⑤ 朴素贝叶斯 (Naive Bayes)

- 贝叶斯的原理是什么？
- 朴素贝叶斯公式是怎么得到的，推导一下？
- 贝叶斯公式
- 先验概率和后验如何用到朴素贝叶斯上的？

⑥ 决策树 (DT)

- 你了解决策树么？讲一下？
- 决策树是怎么样进行划分的，决策树的损失函数？
- DNN 和树模型，哪个用于输入是类别特征时效果更好，为什么？
- 树模型如何输出特征重要性得分？物理含义是什么？

2.5.1.4 无监督学习-聚类方面

- 介绍一下 kmeans 算法，然后问在分布式的环境下怎么实现？
- 说说 k-means 算法和高斯混合模型的关系？
- kmeans, K 值选择，初始点？
- k-means 的数据应该符合什么条件
- 说说高斯混合模型以及参数估计过程？

2.5.2 手推算法及代码

- 写一下逻辑回归的损失函数，并且推导一下权重更新公式？
- 推导 LR、LR 损失函数是什么，数学含义是什么？
- 写线性回归的损失函数，并推导权重更新公式。
- 手推 XGBoost

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

- 神经网络有哪些常用的损失函数？

2.6.2 激活函数方面

- RELU 有什么缺点？
- 推导 softmax 的梯度，和 tanh 的梯度？
- 反向传播梯度推导？（relu、sigmoid）

2.6.3 网络优化梯度下降方面

- 如果不用梯度下降优化，还能怎么优化，我说了一些优化算法，他说牛顿法怎么具体更新参数的。
- 说说最优化方法有哪些，以及具体解释一下 adam 解决了什么问题（从动量以及学习因子自适应的角度解释了一下）
- 具体解释 adam，二阶矩是什么？为什么要用二阶矩？
- 剃度优化 bfgs 的原理
- 写 sgd 和 momentum 优化的方法？
- lda 调参中的参数是什么， α ， β 怎么设置？

2.6.4 正则化方面

- 正则化有哪些方法？详细介绍下 L1 和 L2？
- 正则化与 bias, variance 的关系？
- 说说正则化技术，为什么 L1 正则化可以保持参数的稀疏性？（我回答了两个角度，一个是画约束图，另一个是 L1 正则化本质上是假设样本的先验分布服从拉普拉斯分布，画出拉普拉斯分布曲线解释），然后又推了一下为什么 L1 正则化本质上是假设样本服从拉普拉斯分布。
- 从两种数学角度解释 L1、L2（一个是画图 一个是数值角度）

2.6.5 过拟合&欠拟合方面

- 过拟合和欠拟合的原因及解决方式？

3 快手面经涉及项目知识点

第三节
快手面经
项目知识点
(整理: 江大白)
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

- RCNN、Fast-RCNN、Faster-RCNN 讲一下？
- 对 anchor 的理解？
- 手写 NMS 代码

3.1.2 图像分类

- 介绍图像分类用到的 Loss
- 一般做分类我们都用交叉熵损失，交叉熵损失是否可以替换为 MSE 损失呢？

3.1.3 人脸识别

- center loss 作用
- 小论文中改进了 center loss，介绍 center loss
- 开放问：为什么 center loss 没有得到广泛的应用？
- 问到在处理上万的类别的分类任务上，centerloss 能否使用，arcface 有什么问题？
- 人脸 arcface 中的 arc 为啥比 cosface 那些好，为啥比 triplet 好？

3.1.4 音频算法

- 音频部分基础知识：

(1) MFCC、FBANK 提取过程、区别。

(2) 帧移，为什么选择这个范围

采样率、带宽、奈奎斯特定理等

- 问几个现实场景中的问题，说几个解决方案或思路。

(1) 说话人识别注册语音少怎么解决：数据扩充、自适应等。

(2) 应对噪声场景：数据扩充、降噪、GAN 网络、说话人分离、DNN VAD 等。

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 的 Position Embedding 是怎么处理的？
- Bert 了解么，position embedding 是什么，哪种效果好？
- 介绍一下 BERT？介绍一下两个训练任务？介绍一下如何 fine-tune？如何在 bert 后接网络？为什么要比直接输出好？

② Attention

- attention 计算时为什么要除根 dk？

③ Word2vec

- 讲一下 word2vec 原理？
- word2vec 具体的细节，损失函数？
- word2vec 原理？skipgram 和 cbow 详细说一下？哪个用的多？skipgram 为什么在解决罕见词方面更好？训练中两种优化方式介绍一下？为什么能优化？
- 介绍一下 Glove？和 word2vec 的区别？

④ 其他

- tfidf 公式写一下
- 评估两个句子的相似度，有什么方法

3.3 强化学习

- 生成式模型与判别式模型的区别？
- 说一说基于值函数和基于策略梯度的 RL 算法的区别，什么时候用值函数 RL 算法合适，什么时候用基于策略梯度的合适？
- 说说 onpolicy 和 offpolicy 的区别，并分别举几个代表性算法？

- 写一写 q 函数的原始定义，并说说 q 函数的意义？
- 你对 GAN 有了解是吧，写写 GAN 的损失函数吧，并解释一下 G 和 D 的训练过程。
- 写一下 RL 中 Q 函数的表达式，并说说 Q 函数的意义。
- 你在项目中用了 DDPG，说说 DDPG 算法吧，（我说的时候可能表述的不够清楚，面试官让我画出几个网络之间的输入输出关系）。面试官接着问，这是 offpolicy 算法吧，你先解释下 offpolicy，并说一下你在项目中怎么实现的。（我回答的核心是在 π 策略上加了噪声，形成采集样本策略）面试官说加的噪声少了会怎么样，加的大了会怎么样，有什么影响）
- 你对 DQN 了解吗，对 DQN 的改进算法了解多少？说说它对 Q-learning 的改进在哪些方面？
- 项目是关于 DQN 和 DDQN 的，后来问了 DDQN 能从根本上解决 DQN 的高估现象吗？

3.4 机器学习方面

3.4.1 推荐系统

3.4.1.1 讲解原理

- 讲一下 DeepFM 的原理？
- 协同过滤，userCF、itemCF 讲一下？
- 推荐方法用了哪些？冷启动怎么实现
- CTR 损失函数及意义？CTR 评价指标？
- youtube 召回模型
- deepfm 模型
- 推荐系统的各个环节（召回、精排、重排）都是干啥的，你有什么理解？
- 如果给你一个 DNN 或逻辑回归模型，怎么输出特征的重要性得分？
- 广告算法岗：
 - (1) 了解计算广告的常见收费方式吗？eCPM 是什么含义？
 - (2) 了解竞价机制吗？解释一下广义第二竞价机制？

(3) 为什么要用广义第二竞价机制呢？（优点是什么）

(4) 了解哪些常见的广告形式？

● 聊到推荐系统，我说了 FM 和 FFM，面试官问我 FFM 的第一个 F 代表什么，我回答说 Field。

面试官让我解释一下为什么要用 FFM？

● 讲一下 FM、FFM、DeepFM 的流程？

3.4.1.2 手写代码

● fm 的公式推导，怎么写成矩阵形式

● 手推 FM/FFM 公式

4 数据结构与算法分析相关知识

第四节
快手面经
数据结构与算法分析
(整理: 江大白)
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

● 给定一个数组 [a,b,b,c,c,c,b,a,b,c]，按照字母序排序 ‘a’ , ‘b’ , ‘c’ ，变成 [a,a,b,b,b,b,c,c,c,c] ？

● 循环数组找最小值

● 旋转数组的查找

● 合并有序数组 (C++)

● 小和问题：在一个数组中，每一个数左边比当前数小的数累加起来，叫做这个数组的小和。
求一个数组的小和。

例子：

[1,3,4,2,5]

1 左边比 1 小的数，没有；

3 左边比 3 小的数，1；

4 左边比 4 小的数，1、3；

2 左边比 2 小的数，1；

5 左边比 5 小的数，1、3、4、2；

所以小和为 $1+1+3+1+1+3+4+2=16$

● AABBCDDEE 给一个数组，除一个数组外均为重复出现，要求找出单独的字母，时间复杂度尽可能低？

● 以时间复杂度 $O(n)$ 从长度为 n 的数组中找出同时满足下面两个条件的所有元素：

(1) 该元素比放在它前面的所有元素都大；

(2) 该元素比放在它后面的所有元素都小

● 给一个长度为 N 的数组 里面有 $1-N+1$ 这些 unique 的数字，其中少了一个 怎么找出少的这个，不允许用额外 space

4.1.1.2 链表

● 链表反转

● 判断一个链表是否有环，有环的话输出环的第一个节点，没有的话输出空。

● 写一个合并 K 个排序链表的代码？

● 合并两个排序链表并删除重复值？

● 做了将 K 个链表排成一个排序好的单链表

4.1.1.3 字符串

● 删除字符串中多余空格，连续空格都变成一个，要求时间空间复杂度尽可能小

● 找出字符串中最长的不含重复字符的子串长度，用双指针轻松解决

- 字符串全排列

4.1.2 树

4.1.2.1 二叉树

- 判断一个二叉树是否为二叉搜索树。优化空间
- 一颗二叉树，找到找到叶子节点的值等于给定值的那条路径？
- 非递归实现二叉树翻转、分割子树求乘积 max？
- 求二叉树从根节点到叶子节点路径等和等于 target 的路径
- 二叉树的友视图
- 翻转二叉树
- 层次构建二叉树，然后先序遍历输出
- 层序遍历二叉树
- 二叉树的蛇形遍历
- 二叉树 Z 形遍历
- 二叉树非递归前序遍历
- 求树中两个节点的最近公共父节点，时间复杂度
- 给一个二叉搜索树，和一个区间，删掉不在区间内的节点？

4.1.2.2 堆

- 最大堆实现
- 堆和栈的区别，应用

4.1.3 排序

- 归并排序(只写了 $O(n)$ 空间的), $O(1)$ 空间能做？
- 归并排序的平均复杂度是多少？最坏复杂度是多少？
- 最大的 k 个数

- 求 $\text{top}(k)$ return $\text{sorted}(A)[::-1][:k]$?
- 链表快排、为什么数组排序都用快排不用归并?
- 手写快排
- 快排的平均时间复杂度，最坏情况是什么，复杂度多少?

4.1.4 搜索

- 快排 dfs

4.2 算法思想实战及智力题

- 从文件中读取 n 个数，求最大的 K 个数（约为 K 个即可）写入新文件；要求： n, k 都很大，无法直接装入内存，空间复杂度 $O(1)$ ，尽量减少 IO 次数；思路：利用概率密度
- 给几种硬币，凑一个数，求最少硬币数
- 最长递增子序列
- dfs 全排列
- 有重复的数组，两个数相加= n ，求两个下标，列出所有可能情况
- 连续最长子序列
- 二分查找：有序递增不重复序列，找出第一个缺失的元素，加强版：如果可以重复呢？
- 最大值和最小值的差小于或者等于 num 的子数组数量
- 实现编辑距离， $O(n)$ 空间复杂度

4.3 其他方面

4.3.1 数论

- 凸函数是什么，有什么良好的性质？极值是什么？
- 解释一下什么是凸函数？（我回答了 Hessian 矩阵半正定就行，）面试官接着问如果函数不可导怎么判断？
- 最小二乘法原理？

- 极大似然函数和极大后验函数是啥？
- 极大似然估计，条矩阵问题： $AX=Y$ ， A 列数大于行数，为什么会有无穷个解，有无穷个解的条件，如何得到一个最优解？
- 如果有 2 组独立同分布的数据，但模型分布不一样，怎么解决？EM 算法，混合高斯模型？
- K 个独立高斯同分布随机变量的结果是？
- $ax=b$ ，求 x 的方法。（求逆矩阵，如果不可逆怎么办。什么情况下可逆。）
- 求 a, b 独立且服从 0-1 均匀分布，求 $|a-b|$ 的期望
- 两个独立同分布的骰子，求骰子数积的期望？
- 扔硬币，连续扔出 2 次的最大期望值，连续扔出两个正就停止
- 求两个数的汉明距离？
- 求 $y = \sqrt{x}$

4.3.2 概率分析

- 打可乐,机器坏了每次打 0-1 杯,如果 $< 1/2$ 再打一次,最多打两次……最后 $< 2/3$ 就投诉,问最后投诉的概率？
- 写一个随机函数发生器，随机产生 (1,2,3,4) 四个数，当采集了无穷多数以后，产生的数概率服从 (0.1,0.2,0.3,0.4) 分布？
- 54 张牌抽出四张，抽出为 1234 的概率
- 三个人斗地主，出现王炸的概率
- 一个孤岛重男轻女，直到生出男孩为止，男女出生率 1 比 1，初始比例也是 1 比 1，问最终的男女比例？
- 某疾病发病概率 $1/1000$ ，患者有 95% 的概率检测出患病，健康者有 5% 的概率被误诊，问若一个人被检测出患病，实际患病概率是多少？
- 30 分钟内看到一辆车的概率是 0.95，求 10 分钟内看到一辆车的概率
- 圆上任取 3 点，得到一个锐角三角形的概率

4.3.3 矩阵运算

- 说说矩阵分析里面特征值和特征向量的意义？
- 矩阵问题： $AX=Y$ ，A 列数大于行数，为什么会有无穷个解，有无穷个解的条件，如何得到一个最优解？
- Leetcode 困难难度:一个矩阵，1 能走，0 不能走，可以上下左右走，问最短路径，时间复杂度(nm),空间复杂度(nm)

4.3.4 其他

- K 个独立高斯同分布随机变量的结果是？
- 最大连续子序列和，并返回开始和结束的位置
- 两个均匀分布相加是什么分布

4.4 Leetcode&剑指 offer 原题

- Leetcode 23
- LeetCode 42：接雨水
- Leetcode 103
- Leetcode 113
- Leetcode 199
- Leetcode 206
- Leetcode 300
- Leetcode 322
- Leetcode 542
- Leetcode 原题：两数之和
- Leetcode 原题：用栈实现队列
- Leetcode 原题：最长回文子串

5 编程高频问题：Python&C/C++方面

第五节
快手面经
编程高频问题
(整理: 江大白)
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 网络框架方面

5.1.1.1 Pytorch 相关

- Pytorch 实现把一个 Tensor 中大于 0 的数字都置为 0

5.1.1.2 Tensorflow 相关

- tensorflow 原理，keras 和他的区别？

5.1.2 基础知识

5.1.2.1 线程相关

- 讲一下 python 的多线程和多进程

5.1.2.2 讲解应用

- Python 引入其他库，底层实现是怎么样？
- Python 的 sort，如果两个次数相同，sort 会改变他们的先后顺序吗？
- 判断字符、取字符、输出二维数据的行列

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 内存相关

- 堆和栈存储
- 两个栈实现队列

5.2.1.2 讲解原理

- 多态（虚函数）
- 空类的大小
- 静态函数，静态变量
- 类里面的变量+static 有什么作用？
- C：指针占用内存大小、strlen 和 sizeof 的区别、堆栈的区别、double 型与 0 怎么比较
- C++：STL 容器、虚函数、类的一些知识
- 什么是深拷贝浅拷贝？

5.2.1.3 讲解应用

- vector 和 list 底层实现？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
快手面经
操作系统高频问题
(整理：江大白)
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

6.2.1 常用命令

- 怎么设置环境变量: .bashrc, 然后 source 一下?
- 统计当前文件夹下面.jpg 文件的数量: `ls -l | grep '\.jpg' | wc -l`
- 计算当前文件夹的大小: `du -sh`

6.2.2 其他问题

- 数据挖掘算法岗问到的问题:
 - (1) spark 执行机制
 - (2) spark sql 写两个表字段相同 id 和 ip, 统计两个表中不一样的 ip 的出现次数排序

7 技术&产品&开放性问题

7.1 技术方面

- 介绍天池比赛 (讲训练验证数据划分以及原因、残差特征、lgb 和 xgb 的选用、具体的模型融合方法、AUC、FM 和 LR 的区别、LR 的损失函数推导、交叉熵公式、AUC 高一定预测的转化率准确吗?)
- 介绍在腾讯实习的项目 (讲 DeepFM、讲 word2vec 原理、word2vec 中的优化: 哈夫曼树 and 负采样 and 频繁下采样、讲协同过滤、AUC、连续值分桶)
- 如何判断一张图片含有什么噪声?
- 场景题: 诈骗项目, 想选出 topk 的诈骗类 (没有标签)
 主要类别有电商诈骗, 虚拟交易诈骗, 婚恋诈骗, 主播诈骗等
 数据有: 举报人 ID, 被举报人 ID, 私信数据, 评论数据, 浏览记录
- 样本不均衡的处理方法
 权重调整, 采样, 对样本不均衡不敏感的指标,

面试官补充：对样本不均衡不敏感的模型，比如 SVM，XGB

然后讨论了一波为什么 XGB 对样本不均衡不敏感？

某个项目为什么效果不好，有何反思，未来如果再做你会怎样改进？

● 场景设计题：现在我们要给快手上的视频用半监督方法打标签，target 是印度风的视频，已知印度快手上传的视频中印度风 bgm 的比例远高于其他国家，现在给你视频的 bgm 和上传国家两个特征，设计一个方法给这些视频打上是否是印度风的标签。个人感觉这个有点像半监督召回问题。

● 如何理解数据挖掘/机器学习/数据科学之间的关系？

● 在实际的业务数据中，现在有一个文件不断保存业务过来的流式数据，如何从文件中等可能的取出想要的某个数据，回答了挺久但好像都不是他想要的回答。

● 你如何解决一段文本去找相关图片的任务？

答：通过文本摘要得到关键字，通过搜索引擎去搜索相关图片，然后利用图片描述，与相关文本做相似度计算，取其 TopN

7.2 产品方面

● 风控部门，如何识别那些经过加工方式上传的恶意视频

15|小米算法岗武功秘籍

1 小米面经汇总资料

第一节
小米面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 小米面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 小米面经-53 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【数据挖掘实习生】

(2) 全职岗位类

【NLP 算法工程师】、【计算机视觉算法工程师】、【视觉图像算法工程师】、

【推荐算法工程师】、【未来星数据算法研究员】、【信号处理算法工程师】、【相机算法工程师】

1.3 面试流程时间安排

小米面试流程-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	深挖项目，重视基础
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	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：

- 博士只有两轮，硕士可能会有三轮技术面

1.4 小米面试心得汇总

★ 项目扣的很细，问的很深入，基本上能聊很久。

★ 面试官都非常不错，属于那种面试风格非常温和然后态度非常友好，遇到不会的时候会提示你而且会延申一下你的认知，总之面试体验非常好。

★ 整个面试下来，感觉问的基础题偏多，机器学习的内容偏多，基本没怎么聊深度学习相关的事情。工程方面的问题也有涉及，感觉应该是推荐系统早期的建设阶段，更多的工作内容偏向于工程落地实现。

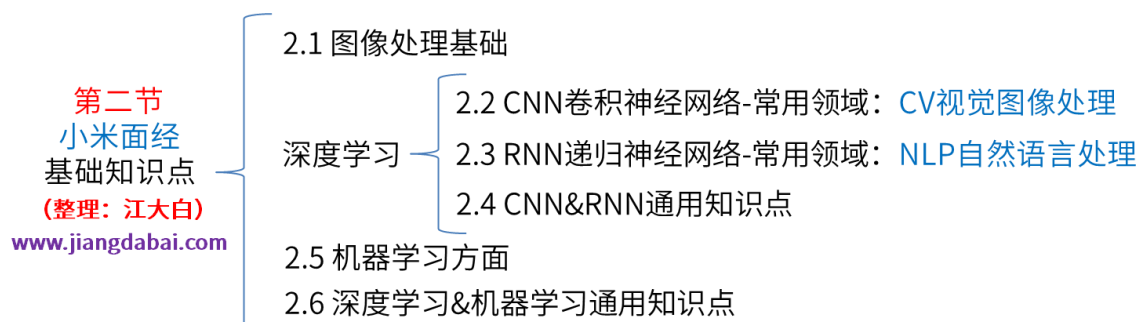
★ 面试官很专业，基本上都是抠细节问问题，需要对领域有一个比较全面的认知。

★ 重点还是问项目吧，项目问的巨细，基本上每个点都会深挖，有些根据项目拓展出来的知识点一定要搞清楚呀，还有自己的项目，一定要自己挖透测，原理，公式，为什么这个方法等等。

★ 小米面试时，先自我介绍，我的习惯是经历简单介绍一下，然后自然转向准备最充分的一个项目开始详细讲，面试官感兴趣的话最好，不感兴趣的话会直接打断的。主要介绍了项目的背景，难点和解决方案，面试官关心的点主要集中在问题抽象和损失函数，讲清楚为什么这么

做，项目大概聊了半小时左右。

2 小米面经涉及基础知识点



2.1 图像处理基础

2.1.1 讲解相关原理

- 讲一下 SIFT 原理
- 介绍 LBP
- 高斯卷积核如何优化，效率提升多少？
- 图像特征检测，边缘检测用过哪些？

2.1.2 手写算法代码

- 手写中值滤波
- 积分图均值滤波如何实现

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 卷积方面

- 除了常用的卷积之外，知道多少种卷积，介绍一下？
- 空洞卷积的原理？
- MobileNetV2 中 1x1 卷积作用？

- 分组卷积优点（减少参数量，从参数计算公式分析）
- dropout 的原理？
- dropout 在测试时用的全部神经元，会不会有偏差，有偏差怎么办？
- 卷积神经网络和全连接神经网络有啥区别？都有哪些层？

2.2.1.2 池化方面

- 最大池化和均值池化适用的场景？

2.2.1.3 网络结构方面

- 熟悉那种神经网络？卷积神经网络？那你说说对卷积神经网络的理解？

2.2.1.4 其他方面

- 梯度消失的解决办法？
- 测量网络大小的三种指标，参数量，flop 计算量，乘加树？（有大佬知道什么是乘加树吗？被问的时候一脸懵逼。）
- BN 的原理，BN 的批处理处理的是什么？
- 白化和 BN 的区别？

2.2.2 数学计算

- conv 和 dwconv 的参数量和 FLOPs 计算公式？
- MobileNetV2 module 的参数量和 FLOPs 计算？

2.2.3 公式推导

- Softmax 原理，推导？

2.2.4 激活函数类

- Softmax 的损失函数？

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- LSTM 的原理？
- LSTM 和 RNN 区别？
- LSTM 每个门的公式还会写吗？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 超参数选择方法，参数类别多时，如何优化？

2.4.2 模型评价

- AUC, ROC 含义？
- 召回率是什么？
- AUC 公式
- PR 和 AUC 的区别？

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

- 数据量大，聚类算法为什么要进行降维？（计算量，维数灾难）
- KSVD 原理？

② 特征选择

- 特征选择的常用方法？
- 如何判断哪些特征是重要的，特征重要性如何计算？step-wise。

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）

- bagging 和 boosting 的区别？
- 随机森林和 bagging 算法有哪些区别？
- GBDT 和 RF，算法思路，优缺点？
- 集成模型一定比简单模型好吗？为什么？
- 随机森林和 gbdt 区别？rf 和 gbdt 哪个容易过拟合？
- xgboost 的默认深度？
- 为什么集成学习里用决策树模型比 LR 这些弱模型的区别？

A. 基于 bagging：随机森林

- 介绍下随机森林？
- 随机森林的原理？
- 面试官问了下随机森林代码调参应该调哪些参数，如何调参？
- 查看结果的指标都有哪些？这些指标的具体含义？

B. 基于 boosting：Adaboost、GBDT、XGBoost

- gbdt, xgboost 模型的比较？gbdt 如何分裂？
- LightGBM 和 xgboost 区别？lightgbm 为什么 更快？
- LR 和 GBDT 的区别？
- lightGBM, XGBoost, GBDT 怎么分裂，怎么找最优分裂点？（xgboost 如何选择最优分割点）
- lightGBM 怎么调参的，网格搜索有没有用过，过拟合先调节哪个参？

- lightgbm 创建每棵树时速度是均匀的吗？（不会）
- lightgbm 训练时和特征数目更相关还是样本数目更相关（不会）
- 连续特征值在 lightgbm 中如何找到分界点？
- lightgbm 该如何调参：找最佳组合-怎么找的-暴力找的-网格搜索有没有用过？
- lightgbm 如果过拟合了，首先调哪几个参数？
- gbdtd 的梯度为什么能代替残差？
- gbdtd 的损失函数是什么？
- gbdtd 如果有 100 棵树，每棵树的输出是什么（我觉得是拟合的上一步输出与目标值的残差，但是具体公式不清楚）
- 介绍 xgb
- (1) gbdtd 和 xgb 的区别(居然没有问 lgb)
- (2) 怎么选最优分裂节点，怎么加速，预排序有什么作用，怎么分箱，等宽还是等深
- (3) 怎么处理缺失值的，预测时候缺失值怎么办

② 逻辑回归 LR

- LR 损失函数公式？
- LR 与 SVM 的区别是什么？并深挖，如 LR 和 SVM 的推导可不可以用梯度下降？
- 推导 LR，写出 loss 和梯度

③ SVM（支持向量机）

- SVM 的原理说一下？
- SVM 的损失函数是啥？
- SVM 的目标函数，为什么能用拉格朗日乘子法讲原始最优化问题转化为极大极小问题，数学原理是什么（不会，面试官也说这个问题有点难了）

④ 朴素贝叶斯（Naive Bayes）

- 朴素贝叶斯为什么“朴素”？（属性相互独立）

⑤ 决策树（DT）

- 信息增益和信息增益比的区别？
- 描述决策树，如何选特征，怎么划分，怎么剪枝，介绍信息增益？

2.5.1.4 无监督学习-聚类方面

- KMeans 怎么初始化？
- 讲了下 DBSCAN 和 K-means 的区别？
- 讲一下 DBSCA？

2.5.2 手推算法及代码

2.5.2.1 手推公式

- 手推 GBDT 公式（要体现找分裂点的过程）
- 手推逻辑回归？
- 手推 xgboost？

2.5.2.2 手写代码

- 手写 kmeans？

2.6 深度学习&机器学习面经通用知识点

2.6.1 激活函数方面

- 深度学习常见的激活函数以及它们的特点？

2.6.2 网络优化梯度下降方面

- 深度学习常见优化方法有哪些？
- 介绍一下梯度下降的原理？

2.6.3 正则化方面

- L1 和 L2 的区别？
- L1 为什么能稀疏矩阵，L2 为什么不能，L2 为什么能解决过拟合？

2.6.4 过拟合&欠拟合方面

- 什么是过拟合？过拟合的解决办法，以及在你的项目中怎么用的？
- 深度学习中过拟合处理，BN 和 dropout。传统机器学习中呢？

2.6.5 其他方面

- 为什么要进行归一化？优点？（量纲，等高线）

3 小米面经涉及项目知识点

第三节
小米面经
项目知识点
(整理: 江大白)
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

3.1.1.1 讲解原理

- 目标检测中，对该任务贡献最多或者改进最大的模块或者网络设计思路？
- 讲一下 anchor-based 和 anchor-free 两者的优缺点？
- 近年来，一些新的目标检测的 backbone 有哪些，各有什么特点？

3.1.1.2 手写代码

- 手写 nms

3.1.2 图像分割

- 目标追踪的常见方法有哪些，包括深度方法和传统方法？各自的应用场景和区别？

3.1.3 图像分类

- 问项目，知道哪些分类的网络；项目用的什么数据集？

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 CRF

- 了解 CRF 吗？

② Word2vec

- 介绍一下 word2vec？
- node2vec 原理？
- word2vec 的两种模型了解吗？工程上的优化了解吗？

③ 其他

- 诸如 ID 类的特征如何处理，编码方式 one-hot 还是其他的，高维时？什么样才算高维，有没有界定？
- CBOW 的网络结构写一下

3.3 强化学习

- SRGAN 和 ESRGAN 的区别？

3.4 机器学习方面

3.4.1 推荐系统

- 介绍一下 CTR 预估里面的模型？
- 画一下 DeepFM 的示意图？然后解释一下这些的区别？
- 系统设计需要注意什么，如何设计一个系统，系统性能如何评估，需要考虑哪些指标（考察点应该是线上的系统了，指标比如内存使用率，qps，99 39 49 时间之类的）

3.5 数据挖掘

- Hive 与 Spark 区别，为什么使用 Spark 而不使用 Hive?
- Spark 常使用的参数，当读取 HDFS 时有一个文件过大，读取报错，添加什么参数可以解决，小文件使用的参数，如何解决数据倾斜，主要想问源码。
- 常使用的 spark 参数，spark 源码中一些 groupbykey, aggregateByKey, groupbykey 源码及其区别
- 怎么保证在 spark 运行阶段动态分配所需的资源，需要添加什么参数，这里问题太多，很多新版本参数我根本不会
- Flink 与 Spark 的区别，Flink 的源码相关，各自的容错机制，基本思想，一次语义，源码如何实现
- Hadoop, Spark 都使 Yarn，介绍一下 YARN，执行过程
- 业务相关问题：验证一条微博突然爆起来，应该监控哪些指标？

3.6 信号处理

- 针对项目具体发问，包括你的算法主要实现思路、和其他算法相比优势在哪？创新点在哪？
- 了解语音信号降噪的常用方法吗？因为我做的都是雷达相关的，直接说不了解了
- 问了一下数字信号处理、信号与系统的知识
- 滤波器了解吗？项目里有用过吗？怎么设计一个滤波器？
- STFT 中窗函数的选择对时频特性的影响？加矩形窗和汉明窗有什么不一样？
- STFT 交叠？
- 采样率是多少？1000Hz；低通抗混叠滤波器截止频率设置的多少？
- 对模拟滤波器了解吗？
- 说一下巴特沃斯滤波器的极点分布
- 短时傅里叶变换了解吗？
- 短时傅里叶变换加窗函数的作用是什么？

- 在不知道阶数的情况下，怎么得到 IIR 滤波器的频率响应曲线？
- 已知 IIR 滤波器系统函数，设计 FIR 滤波器？
- 怎么得到 FIR 滤波器的频响曲线？

4 数据结构与算法分析相关知识

第四节
小米面经
数据结构与算法分析
(整理: 江大白)
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

- 排序好的数组，截断后，后部接到前部的前方，构建成新的数组，如何找出特定元素？
- 循环有序数组的查找 (C++) 二分查找变体
- m 个数组，每个数组里有 q 个数，找出其中第 k 大的数
- 升序旋转数组找最小
- 两个有序数组，求其中位数，然后改进时间复杂度？
- 两个数组合并求中位数
- 子数组最大和？
- 数组中出现次数最多的 K 个数
- 从升序的数组中找出出现次数大于等于数组长度一半的数字？
- 包含重复数字的无序数组，找到所有加和等于 $target$ 的索引对？
- 给定一个包含从 $0, 1, 2, \dots, n$ 中取出的 n 个不同数字的数组，找到数组中缺少的数字？
- 一个长度为 n 的数组，求其中出现次数大于 $n/2$ 的元素，要求线性时间复杂度，恒定空间

复杂度?

- $m \times n$ 的二维数组，只能往右或者往下，找最短路径?

4.1.1.2 链表

- 链表反转
- 链表对折 1 2 3 4 5 变成 1 5 2 4 3

再拆解一下题目，(灵活)

(1) 找到链表的中点

(2) 翻转后半段链表

(3) 合并两个链表

- 判断两个单链表是否相交
- 两个链表求交点
- 链表 1->2->3->4 反转为 1->4->2->3
- 单链表倒数第 k 个节点
- 用两种方法复制复杂链表
- 求取单链表的中间? (①遍历求表长，取 $n/2$ ②快慢指针)

4.1.1.3 字符串

- 字符串最长不重复子串 (哈希表，思路和买卖股票一样)
- 给出一个字符串，写出该字符串的全部排列组合?
- 判断字符串中括号是否对应?

4.1.2 树

4.1.2.1 二叉树

- 二叉树公共父节点
- 求二叉树的右视图?

- 用非递归的方式实现一个二叉树的删除操作,引申问题 $O(1)$ 的空间复杂度怎么解决?
- 树的层序便利
- 二叉树最深、最浅深度
- 二叉树中序遍历非递归 (C++)
- 宽度遍历二叉树
- N 个结点的二叉搜索树有多少种组成方式?

4.1.2.2 堆

- 堆排序

4.1.3 排序

- 写一个归并排序
- topK 的两种思路
- 手写快排序
- 手写堆排序

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 三色旗问题
- 买卖股票的合适时间的算法题
- 括号匹配问题, 检查是否合法? 如果不用栈, 如何解决括号匹配问题?
- 最长回文子串, 动态规划
- 股票题, 一个数组, 问最大收益
- 二分查找轴上索引, $list = [0.0, 1.0, 2.0, 3.0]$, $value = -1$ 返回 0, $value = 0.5$ 返回 0, $value = 1.5$ 返回 1, $value = 3.5$ 返回 3

4.2.2 智力题

- 马匹赛跑：25 匹马，5 个跑道，没有计时器，要找出前三名，最少要比多少场？答案是 7
- 9 枚硬币，8 枚一样重，1 枚比较重，最少称几次能找到最重的那枚？2 次

4.3 其他方面

4.3.1 数论

- 求两个数最小公倍数 (C++)，最小公倍数= $a*b/\text{最大公约数}$ ，最大公约数：辗转相除法？

4.3.2 概率分析

- 极大似然估计的原理？然后给出一个二项分布，让用最大释然估计手推出该分布的参数？
- 一个线段分成三段，求他构成三角形的概率？
- 一个整数数组，其中有几个数字 1，如何等概率抽取 1 个 1？

4.3.3 矩阵运算

- 给定 $n*n$ 矩阵，计算新的 $n*n$ 矩阵。新矩阵的 (i,j) ，是以原矩阵 (i,j) 为中心的 $m*m$ 小矩阵的均值 (m 为奇数)。
- 写一个矩阵旋转 90 度的代码，解释一下坐标变换公式
- 给出一个二维矩阵，顺时针由外层到内层打印该二维矩阵？

4.4 Leetcode&剑指 offer 原题

- Leetcode 166

5 编程高频问题：Python&C/C++方面

第五节
小米面经
编程高频问题
(整理: 江大白)
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 网络框架方面

- 平时深度学习用什么框架？

5.1.2 基础知识

5.1.2.1 讲解原理

- 说一下 Python 中 list 的底层代码怎么实现的吧？

5.1.2.2 讲解应用

- Python 包 都有哪些？

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 线程相关

- 进程与线程，线程共享的资源（堆，全局变量）

5.2.1.2 讲解原理

- 虚函数表（主要就是动态绑定，继承类调用基类虚函数，指针数组，虚函数的指针）
- 基类的析构函数为什么用虚的析构函数？（防止内存泄漏）
- 全局变量存放位置

5.2.1.3 讲解应用

- C++内存管理，预编译

6 操作系统高频问题：数据库&线程&常用命令等

第六节
小米面经
操作系统高频问题
(整理: 江大白)
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

6.1.1 基础问题

6.1.1.1 区别比较

- Left join 跟 join 的区别？
- Join 的话，大表和小表哪个连接在前？

6.1.1.2 讲解原理

- SQL 喜欢考窗口函数

6.2 操作系统方面

- 进程和线程的区别？
- 栈，队列，和堆的区别？

7 技术&产品&开放性问题

7.1 技术方面

- 场景分析题，如何对新闻进行实效性分析，怎么挖特征？
- 如何从大量网页中找出新闻类网页，算法设计？

16|深信服算法岗武功秘籍

1 深信服面经汇总资料

第一节
深信服面经
汇总资料

(整理: 江大白)
www.jiangdabai.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 深信服面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 深信服面经-72 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 全职岗位类

【机器学习工程师】、【NLP 算法工程师】、【大牛计划算法工程师】

1.3 面试流程时间安排

深信服面试流程 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	主要是项目细节及创新点 关注解决问题的思路
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	主要是项目细节及基础知识
第三面	技术Leader面	自我介绍+项目经验+公司发展	/
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。

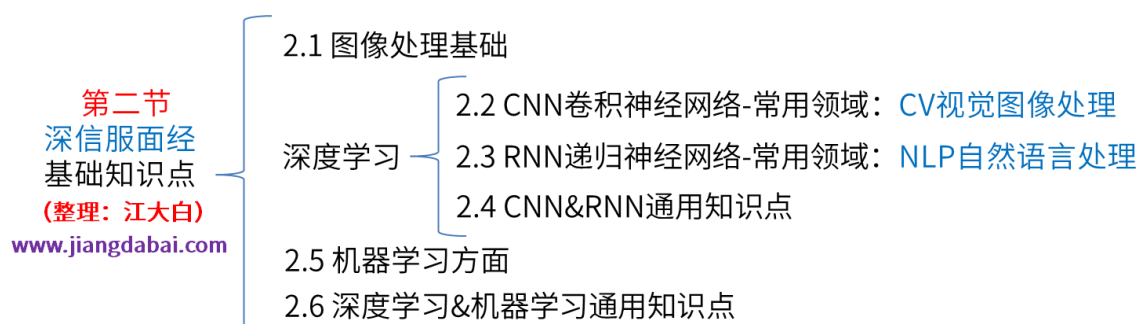
其他注意点：

- 第三面技术 Leader 面有的时候会有

1.4 深信服面试心得汇总

- ★ 比较看重及解决问题的能力以及项目落地开发的能力
- ★ 深信服是我面了这么多公司中感觉面试最难的一家公司，真的很考验技术。
- ★ 很多都是 python 开发及前端、后端开发的，统计的主要还是统计机器学习相关的
- ★ 感觉机器学习问的多一些，深度学习的不多
- ★ 不同公司定义算法工程师不一样，深信服的是做什么，他说是做安全和云计算，所以更多的是会做一些安全工作，比如使用算法进行漏洞检测，攻击检测这些

2 深信服面经涉及基础知识点



2.1 图像处理基础

无

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 卷积方面

- 讲一下 CNN 的原理?
- 对于 dropout 的认识

2.2.1.2 网络结构方面

- 谈一谈 Resnet?

2.2.1.3 其他方面

- 深度学习模型的初始化要注意哪些问题? (权重随机初始化, 避免网络退化, 初始化范围要小, 缩小样本空间, 输入样本要 BN, 防止梯度消失)
- 什么是 BN?
- 是否了解自动调参 auto ml?

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- LSTM 的门结构有哪些？门的输入是什么，输出是什么，怎么在网络里面使用？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 注意力模型的原理？
- 如何解决样本不平衡问题？
- 有一个分类任务有几千个类，应该怎么去训练模型？
- Attention 机制讲一下？

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

- 特征如何降维？
- LDA 的实现方法、LDA 中的奇异值分解矩阵实现？

② 特征选择

- 在项目讲述过程中问了几个问题：（1）特征是什么样的？（2）怎么构建特征工程的？（3）为什么选择这个模型？
- 机器学习中有哪些算法需要进行归一化？

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）

- 介绍一下集成学习，以及你选择的方法原因？

A.基于 bagging：随机森林

- 随机森林原理，如何选择最优分裂点，列抽样为什么可以缓解过拟合？

B.基于 boosting：Adaboost、GDBT、XGBoost

- 为什么要模型融合，模型融合的几种方法，模型融合的优点？（bagging,boosting,stacking,还有我自己项目里面的方法）
- 白样本多，黑样本少，选择哪个模型更适合一些，为什么？如何评价效果（抽样，训练多个分类器，可以采用 bagging，如果 LR 和 SVM 里面选择 SVM,LR 对样本过于敏感，SVM 只处理支持向量）
- xgboost 相对于 GBDT 的优点，为什么会快些、xgboost 如何做并行？
- 讲一下 xgboost（从提升树开始讲，讲了一通）？为什么要二阶展开？xgboost 采样的时候怎么采样的？

② 逻辑回归 LR

- LR 和 SVM 的区别？（处理点、计算方式、损失函数、自带正则等）
- LR 过拟合是什么样的情形，如果样本有很多重复的特征，对于 LR 训练效果有没有影响？
- LR 使用什么损失函数，为什么不用差平方，用差平方与交叉熵差别在哪？

③ 决策树（DT）

- 信息熵的定义？

2.5.1.4 无监督学习-聚类方面

- 问了 kmeans 原理，优缺点，如何不自己设置 k 就能知道 k 取多少？

2.6 深度学习&机器学习面经通用知识点

- 训练的模型过拟合了，怎么办？

- 有一个分类任务有几千个类，应该怎么去训练模型？陷入局部最优值怎么办？

3 深信服面经涉及项目知识点

第三节
深信服面经
项目知识点
(整理: 江大白)
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

- Focal loss F1 怎么解决样本不平衡的问题？

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

- 你觉得 BERT 有哪些可以改进的地方？我说您是说 BERT 的缺点吗？
- 文本分类中，传统机器学习 tfidf+lr/svm 和 Bert 的区别？

② Word2vec

- word2vec 原理？
- word2vec 和 onehot 的区别？

3.3 强化学习

无

3.4 机器学习方面

无

4 数据结构与算法分析相关知识点

第四节
深信服面经
数据结构与算法分析
(整理: 江大白)
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

- 给定一个数组 arr, target, 求数组元素的组合之和=target 的所有可能。(不能重复)
- 数组中无重复的和为 target 的所有组合 (回溯+去重)
- 删除数组中重复的数字, 相对位置不变输出, 且保证输出元素满足从小到大顺序的个数最多?
- 给定一个数组, 求所有重复的数。用字典, 问还有没有其他方法, 提示用 hash。

4.1.1.2 字符串

- 有一篇英文文章, 找前 10 个出现次数最多的字母?

我回答先遍历一遍论文, 简历字母为 key 出现次数为 value 的哈希表! 然后从 values 中找到前 10 个最大的元素! 面试官又问, 怎么找? 我说用最小堆。

- 一个特别特别大的字符串, 怎么统计给定字符集里面的字符在字符串中首次出现的位置, 然后加快效率更好? 我说 map-reduce, 他说更好的呢, 用数据结构实现? 我说 hashtable, 在对应位置上面存首次出现的位置就行。

- 英文文本, 得到前 K 大频率的单词, 如果文本很大怎么办? 写伪代码

4.1.2 树

- 平衡二叉树与数组获取元素谁比较快？（如果给定下标，数组快，给定值，平衡二叉树快，原因时间复杂度）

4.1.3 排序

- 排序算法介绍，它们的复杂度？
- 常使用的排序算法，说一些你熟悉的并口头实现？（说的使快排，复杂度 $O(n\log n)$ ）
- 说一说你知道的时间复杂度为 $O(n\log n)$ 的算法。以快排为例，具体解释一下为什么它的时间复杂度是 $O(n\log n)$ 呢？
- 类似快排的这些算法思想都属于什么算法？（分治算法）请你说一说这一类算法的时间复杂度有什么规律？快排是分成两组 1，是否可以分为更多组呢？为什么？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 场景题目：公司组织出游，每个人时间不同，让给出算法求出一个时间段尽可能满足最多人的要求？
- 利用二分法，判断一个数是否在给定的 list 里，并问时间复杂度？
- 红，绿，蓝三种颜色，n 个各自，相邻不能一样，首位不能一样，求填充方法数？

4.2.2 智力题

- 四个人分别花 1,2,5,8 分钟过桥，桥上只能容纳两个，且过河要手电筒！问最短过河方案！
- 想了一下为 $8+1+5+1+2$ ，面试官说这是第二优的解法，但不是第一优的。
- 打灯过桥的问题，4 个人过桥的时间分别为 1 分钟，3 分钟，7 分钟，9 分钟(具体时间忘记了，但思路不变)。只有一个手电筒，每次只能两个人过桥，过桥时间以两个人中最慢的那个人过桥时间为准，问这四个人全部过桥最少需要多少时间？

4.3 其他方面

4.3.1 数论

- 给定 0-100 个数，再给定一个数，怎么判断这个数是不是重复了

4.3.2 计算几何

- 二维平面有 n 个点，求点 p ，其到所有点距离之和最小？
- 给定坐标系内的一个矩形和一个点，怎么判断该点是否在矩形内部？

4.3.3 概率分析

- 一个公交站 1 分钟内有车经过的概率是 q ，那么 3 分钟内有车经过的概率是多少？
- 如果一个人在公交车站台一分钟内能等到公交车的概率为 p ，那么这个人在三分钟内能等到这辆公交车的概率为多少？(从反面去思考 $1-(1-p)*(1-p)*(1-p)$)

4.3.4 矩阵运算

- 一个 $100*100$ 矩阵，里面所有数都是正数，判断从左上角到右下角是否存在一条路径和为奇数的路径？

4.4 Leetcode&剑指 offer 原题

- Leetcode 65

5 编程高频问题：Python&C/C++方面

第五节
深信服面经
编程高频问题
(整理: 江大白)
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 内存相关

- python 里面线程与进程、进程如何共享内存？
- python 的内存处理机制有哪些？分别介绍一下。
- Python 了解么？说一下内存管理机制

5.1.1.2 区别比较

- Python 中的进程，线程和协程的区别？
- 深拷贝和浅拷贝的区别
- python 中 is 和 == 有什么区别？

5.1.1.3 讲解原理

- python 里面的进程、线程、协程特点-说了进程是资源最小单元，线程分配最小单元 线程相互影响？
- 了解 python 装饰器吗？它的作用？
- python 装饰器知道吗？装饰器的原理是什么？本质是？闭包？
- 判断一个值是否在数组里，用 set 快还是 list 快？
- Python 中常用的排序有哪些？
- Python 中的自动化测试用过么？
- 统计各个函数执行的时间（就是想问装饰器，我不会），python 装饰器
- 如果结构体中，声明了一个 char 型变量，一个 int 型变量，那么这个结构体占多少字节？
- python 中的内存管理知道吗，介绍一下？
- python 中 self 的用法？
- python 中的可变对象和不可变对象有哪些，特点、用法？

5.1.1.4 讲解应用

- python 里常用的数据类型有哪些
- python 字典中有 1000 万条数据，如何取所需要 value 值的 100 万条？

5.1.2 手写代码相关

- 介绍下 lambda

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 区别比较

- size of 和 stl: len 的区别

5.2.1.2 讲解原理

- 结构体：内存对齐

6 操作系统高频问题：数据库&线程&常用命令等

第六节
深信服面经
操作系统高频问题
(整理: 江大白)
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

6.2.1 TCP 协议相关

- tcp /udp 的区别

- 三次握手讲一下？
- 为什么三次握手，为什么四次挥手？

6.2.2 线程和进程相关

6.2.2.1 区别比较

- 进程，线程的区别？
- 深拷贝、浅拷贝

6.2.2.2 讲解原理

- 计算机网络，介绍知道的网络知识？
- linuxs 线程和内存管理

6.2.3 常用命令

- Linux 命令了解么？Linux 中软链接和硬链接的区别？
- Linux 中文件权限是怎么样？

7 技术&产品&开放性问题

7.1 技术方面

- 有一个大的文本序列，求出现次数最多的前 k 个单词？复杂度？太多怎么保存？
- 说一下无监督学习？
- 考虑过读取 10M 大小的 json 文件，内存会占用多少？怎么保存的么？
- 垃圾短信多分类任务(如何分开发票，广告，商铺信息等)，有什么思路
- 如果有 n 种类别(比如新闻类，体育类等)的网站，目前收集到一些网站，及其网站中不良信息的位置，那么新来一个网站，如何判断该网站中是否含有不良信息，若含有，不良信息在哪个位置？
- 如果收集到一些网站的语料，如何判断这些网站中是否有不良信息？

他说如果有一些语料，然后还有一些关键词，如何判断这些语料中是否含有这些关键词？

- 大牛计划岗位：怎么设计一个查重系统？
- 场景题，一张多个人的合影，怎么确定人的位置，以及输入工号得到位置，怎么设计网络等等

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.jiangdabai.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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [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

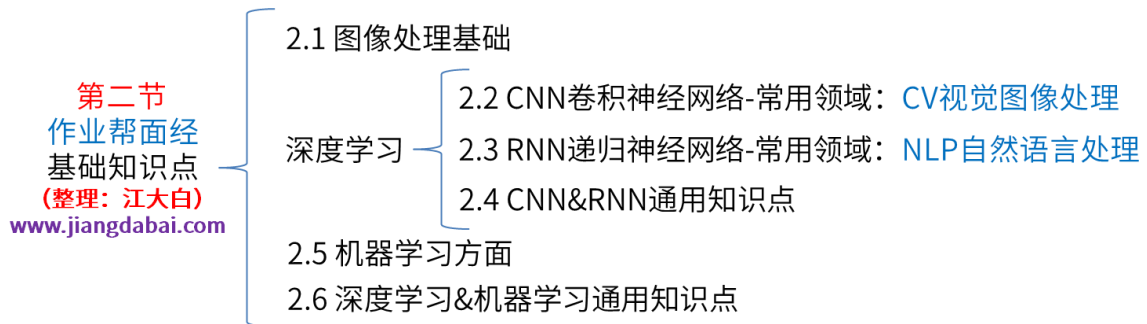
★ 算法会问很多，看重代码能力

★ 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 pooling。

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、GBDT、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 gbdet 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

- 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 networkk 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

- 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 nlog(n)），最长不重复连续子串（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 1×3 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

{

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

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？

18|滴滴算法岗武功秘籍

1 滴滴面经汇总资料

第一节
滴滴面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 滴滴面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 滴滴面经-56 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【滴滴地图事业部算法实习生】、【滴滴算法实习生】、

(2) 全职岗位类

【滴滴推荐算法工程师】、【滴滴大数据算法工程师】、【滴滴新锐算法工程师】、【滴滴机器学习算法工程师】、【滴滴地图事业部机器学习算法工程师】、【滴滴定价策略算法工程师】、【滴滴国际化事业部算法工程师】、【滴滴 R-Lab 算法工程师】、【滴滴金融事业部算法工程师】、【滴滴核

心岗出行算法工程师】、【网约车供需策略算法工程师】、【滴滴出行规划算法工程师】

1.3 面试流程时间安排

滴滴面试流程-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	项目细节和基础知识点
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	项目问题更开放、 偏实际应用
第三面	技术Leader面	自我介绍+项目经验+公司发展	/
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。

1.4 滴滴面试心得汇总

★ 整体感觉滴滴是面试体验很好，面试官是以一种平等的态度来交流技术，即使有时候卡壳也会慢慢提醒。

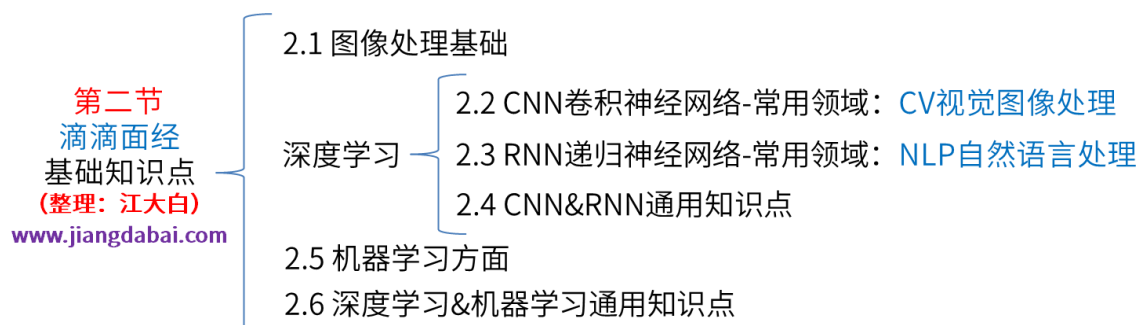
★ 有竞赛的会问竞赛项目，大部分都是围绕比赛项目问的，建议大家对自己的项目了如指掌

★ 主要机器学习为主，基础知识也会问，但不会问的特别深，计算机视觉很少

★ 总体还是不错的。除了一面面试官有点严肃，二三面都是边笑边面。尤其三面面试官，给人印象很不错，有种如沐春风的感觉。

★ 偏向于纯机器学习，我做图像方向的不太适合岗位

2 滴滴面经涉及基础知识点



2.1 图像处理基础

无

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 卷积方面

- CNN 和全连接的区别? (参数共享, 降低运算量)
- 详细讲一下 CNN 怎么卷积多通道?

2.2.1.2 网络结构方面

- 说一下 ResNet 和 DenseNet 的区别? 为什么 DenseNet 比 ResNet 效果好?
- Autoencoder 的原理?

2.2.1.3 其他方面

- 什么是梯度消失和梯度爆炸? 如何解决?
- 梯度消失和梯度爆炸问了一下, 缓解的方法 (Relu, LSTM, BN), 为什么能缓解, BN 的参数?
- Batch normalization 的原理?

2.2.2 公式推导

- 手推两层神经网络(input->FC->sigmoid->FC->sigmoid->output)的反向传播?

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- LSTM 和 RNN 的区别?
- LSTM 原理讲一下? LSTM 的用法/具体应用?
- LSTM 有什么新的改进, 替代? (针对 rnn 来说 通过门电路把连乘转换成了连加, 一定程度上缓解了梯度消失问题, 梯度爆炸可以直接 clip 不需要考虑, 另外就是广度了, lstm 的应用和改进)
- LSTM、CNN、convLSTM 原理, 异同对比?
- LSTM 为什么可以解决梯度弥散问题?
- 连续性变量需要离散化的时候在神经网络里要怎么做?

2.3.2 手绘网络原理

- 手推 LSTM 公式

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 样本类别不均衡你会怎么处理(重采样/Focal loss)?
- Transformer 结构是什么样的, self attention 公式是什么, 你怎么理解 self attention 的?
- 介绍 transformer?

2.4.2 模型评价

- 分类评价指标讲一下?
- 讲一讲常见评价指标, TPR/FPR/ROC/AUC?
- 数据不平衡用 accuracy、PR、ROC 指标会咋样?

- AUC 指标的含义，同时增大减小样本的预测概率（保持顺序不变）会对 AUC 指标产生什么影响？

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

- 如何做特征选择？
- 用了哪些技术做特征工程？（特征筛选+数据预处理）

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）

- 模型融合方法，尝试了几种，都怎样做的？
- boosting 和 bagging 的区别？（bagging 就是投票策略，会降低方差，减小过拟合风险，boosting 是降低偏差的策略，可以描述一下 Adaboost 和 gradient boost 的区别）
- XGB 和随机森林的区别？

A.基于 bagging：随机森林

- 随机森林算法讲一下？
- 随机森林的随机性体现在哪里？
- 说一下随机森林的缺点？
- RF 分裂时特征选择怎么选的，常用的还有什么选法？

- 随即森林是怎么采样的，随机森林的基本分类器有啥？

B. 基于 boosting: Adaboost、GDBT、XGBoost

- 为何使用 xgb, xgb 原理, xgb 并行实现？
- xgboost 与 gbd、RF 的对比？
- gbd 原理？xgboost 源码看过吗？
- xgboost 相对于 gbd 改进了什么？
- 因为模型 xgb 用到了，所以问的很细，xgb 的原理？xgb 和 gbd 的不同？为什么 xgb 用二阶导？xgb 怎么梯度下降的？xgb 的损失函数？xgb 的正则化？
- xgb 和 gbd 区别？（基本是必考题目，从主要的优化点说起，xgb 是二阶泰勒展开，gbd 是一阶，可以类比牛顿法和梯度下降法的区别，牛顿法收敛更快，但是由于更快逼近目标，会增大过拟合风险，因此在目标函数里有一个显示的惩罚项，与叶子节点数和叶子节点的权重有关，来控制模型复杂度；另外还有分裂节点的选择，xgb 怎么选取最优分裂节点，有哪些加速的优化之类的知识）
- lightgbm 相对 xgboost 的改进？
- xgboost 流程说一下？

② K 近邻 (KNN)

- KNN 你知道吧，里面搜索 k 近邻用的数据结构叫 kd 树，你知道吧，说一下 kd 树用在 KNN 的哪个步骤里？
- kd 树的作用是啥？
- kd 树是怎么构建的？它的时间复杂度是多少？为什么 kd 树能够比较快的找到 k 近邻？

③ 逻辑回归 LR

- LR 优缺点？
- 问了下 LR 的损失函数？
- 逻辑回归的损失函数，和极大似然的内在关系？
- LR 做特征重要性分析时，连续值会有影响吗，怎么处理，特征的重要是看 one_hot 后的

sum 还是 mean?

④ SVM (支持向量机)

- SVM 原理, 损失函数是什么, 软间隔怎么做的?
- 讲讲 SVM 吧, SVM 对错分点怎么处理的, SVM 是什么问题?
- SVM 训练完成后得到的模型是什么?

这题我简单说一下, 一般人都会认为得到的模型是那个超平面, 即 $y = w_1x_1 + \dots + w_nx_n$, 我也是这样回答的。但是面试官反问如果是带核函数的模型, 怎么进行预测。因为核函数是计算两个样本之间的高维内积的, 因此只有超平面和另一个待预测样本的话就不知道怎么做预测了。后来面试官跟我说得到的模型其实是支持向量, 预测时将待预测样本与支持向量计算距离。

- SVM 为什么自带泛化性?
- SVM 对偶化的条件是什么? 具体说明每一条, 并且回答为什么满足该条件就可以对偶化?
- SVM 与其他模型相比, 在面对样本不均衡问题是更优还是更劣?
- SVM 的软间隔和硬间隔是什么? 软间隔加入的松弛变量是如何求解出来的?

⑤ 决策树 (DT)

- 决策树有哪几种, 平时用的多的是哪种, 每种决策树有什么区别?
- 决策树如何选择特征? 信息增益怎么计算?
- 决策树的流程? 了解决策树剪枝么?
- 决策树和随机森林的区别?
- cart 和 c4.5 的区别?
- 离散特征经过 onehot 处理后, 直接放入树模型有什么影响?

⑥ 其他

- 介绍下极大似然?
- 遗传算法原理说一下、缺点、常见的编码方式?
- 树模型、NN、NB 的区别和应用场景?
- 让我介绍参加过的京东的比赛(要求按照一定的框架讲, 比如做了什么, 觉得自己做的好的

有说什么，有什么做的不够好的)，在介绍的过程中针对性的提了一些问题，可以说都是一些常考的问题，记录如下：

- (1) 解释下过拟合；
- (2) 数据类别不均衡的处理方法；
- (3) XGBoost 相对于 GBDT 有什么不同；
- (4) 数据记录有多少？特征有多少？训练时间多久？

2.5.1.4 无监督学习-聚类方面

- 介绍常用的聚类算法（KMeans、DBSCAN、Mean Shift）
- K-means 算法的过程讲一下？
- 看你项目中用到了 kmeans，这是分类算法还是聚类算法？请介绍 kmeans 与 knn 算法流程？

2.5.2 手推算法及代码

2.5.2.1 手推公式

- 手写 LR？
- 手推梯度反向传播及 SVM 的过程

2.5.2.2 手写代码

- 手写 K-means 伪代码(计算到中心点距离/判断迭代停止/更新中心点的坐标等核心函数，串一起)

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

- 写了 sigmoid 函数

2.6.2 激活函数方面

- 有哪些常用的激活函数？
- sigmoid 函数做激活函数有哪些缺点？

2.6.3 网络优化梯度下降方面

- 常用的优化算法？
- 牛顿法和梯度下降法的区别？
- 牛顿法有什么优缺点？
- 梯度下降的更新公式，和梯度提升的区别？

2.6.4 正则化方面

- L1 正则和 L2 正则的目的和原理讲一下？
- 问了下 L2 正则会不会影响 adam 的学习率？

2.6.5 过拟合&欠拟合方面

- 你遇到过过拟合么？什么是过拟合？有什么体现？
- 有哪些抑制过拟合的方法？
- 模型发生过拟合了，怎么判断它是数据量不够还是模型复杂了？

2.6.6 其他方面

- 样本不平衡怎么做？

3 滴滴面经涉及项目知识点

第三节
滴滴面经
项目知识点
(整理: 江大白)
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

- Fast-RCNN 和 Faster-RCNN 的区别是啥？
- unet 变体
- anchor free 算法进展

3.1.2 图像分割

- 你说你是做医学图像分类与分割的，说一下你用的经典的图像分割网络(回答得是 UNet 系列)
- 问为啥 UNet 在医学图像分割上面效果比较好？

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 的输入是什么？

② Transformer

- 问了 textcnn、transformer 的细节？

③ CRF

- 为何使用 biLSTM+CRF 的结构？

④ HMM 隐马尔科夫模型

- HMM、MEHMM、CRF 原理以及对比？
- EM 算法的原理，在 HMM 中如何使用的？

⑤ Word2vec

- 介绍 word2vec2 种实现模型、区别以及和 fasttext 的区别？
- word2vec 的原理讲一下？

3.3 强化学习

- 了解强化学习么？
- Q 值和 V 值的联系和含义？
- 强化学习 OFF POLICY 和 ON POLICY 的区别？

3.4 机器学习方面

3.4.1 推荐系统

- 解释了一下 FM、DeepFM 方面推荐的基础模型？
- node2vec 怎么做的？
- deepFM 相对 wide&deep 改进？
- fm, ffm, deepfm 区别？

4 数据结构与算法分析相关知识

第四节
滴滴面经
数据结构与算法分析
(整理: 江大白)
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

- 给定一个无序数组，要求把奇数放在数组前半部分，偶数放在数组后半部分？并且奇数部分和偶数部分均有序。
- 把一个数组分成 k 组，返回所有可能的组合？

- 找到数组中每个数的右边比他大的第一个数，要求时间 $O(n)$?
- 两个有序数组的中位数
- $\log n$ 找出不出现的最小正整数
- 二维数组最短路径，存在障碍的最短路径?
- 旋转数组中的查找
- 有序数组找和目标值相等的最左边的下标
- 两个有序数组，求第 k 大，时间复杂度尽可能低 ($\log(m+n)$)
- 二维数组的旋转打印，比如 1 2 3 8 9 4 7 6 5 从 1 到 9 打印出来?
- 多个有序数组求交集?
- 数组和链表的区别?
- 多个有序数组求并集?
- 数组逆序对
- 给一个数组的字符串，然后求 shape。如：“[[1,2,3],[1,2,3]]”，结果为(2,3)。

4.1.1.2 链表

- 反转链表?
- 单链表的排序，要求时间要达到 $n \cdot \log(n)$
- 两个链表合并
- 链表题，奇数位升序，偶数位降序，重排成升序的链表

4.1.1.3 字符串

- 给一个字符串和一个 k ，要求找到不超过 k 个不同字符的最长子串的长度，on 时间 (LeetCode 上的 medium 或者是偏简单的 hard 题目，用划窗的思路做)
- 求和大于 k 的子串的最小长度?
- 最长公共子串问题
- 给你若干个 list，从每个 list 中选出一个元素，求全排列?

- 动态规划问题，一个长的字符串中找到最长的没有重复字母的字符串？
- 括号匹配问题，给一个不匹配括号串，只包含左括号和右括号，这个串去掉一个字符可以变成匹配串，找到所有可能的匹配串

4.1.2 树

4.1.2.1 二叉树

- 求树深，时间复杂度（LeetCode easy，时间复杂度是 $O(N)$ ，因为要扫一遍所有的节点）
- 给出二叉树的先序遍历和中序遍历，输出后序遍历？
- 之字打印二叉树

4.1.2.2 堆

- 堆排序算法过程？

4.1.3 排序

- 归并排序求逆序对个数，并记录每个元素对应的逆序对个数
- topk
- 非递归快排
- 手写了一个快排
- 实现一个堆排序

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 一亿个结点，每个结点有个 score 和一个 class，按照 score 大小要把排名前 10% 的结点 class 变为 0, 10~20% 的变为 1，以此类推？
- 有 m 件物品，每件物品分别有不同的重量是一个 m 大小的数组， n 个背包 1~ n 编号，每个背包的容量为 T ，把物品往背包里面装，一个背包在容量运行的范围内可以装多个物品，但是装背包必须按照背包的编号来，并且开始往第二个背包里面装物品之后就不能再往前面的背

包里面装了，问最多可以装多少件物品。（类似买卖 k 次股票的问题）

- 数三减一，就是小朋友抱成圈，数到三删除一个人，考代码实现能力
- 国王挖金矿的变体
- 最长递增子序列
- 参数量估计：绳子剪成 m 段，最大乘积
- 因式分解
- 判断年份是否为闰年
- 开三次方根。其实跟开根号的题类似。牛顿法可能要推一下公式，可以用二分法。

4.2.2 智力题

- 一条路上，总长度为 n ，有长度为 1、2、3、5 的砖，问总共有多少种铺法？DP 求解。
- 1000 桶水有一桶有毒，一头猪喝了有毒的水之后过 15 分钟会死掉，问最少需要多少头猪来找到有毒的水，在一个小时之内？

4.3 其他方面

4.3.1 数论

- 讲解泰勒级数，傅立叶变换，复变函数（因为主管是学数学出身）
- 判断质数？

4.3.2 概率分析

- 2 人抛硬币，正面赢，反面让另一个人抛，求先抛的人获胜的概率？
- 抛硬币，连续两次正面向上为止，求次数期望
- 三个门，门后一个有礼物，你选了一个，帮你排除了一个没有礼物的门，问你换不换：换，假设第一门有礼物，计算换和不换拿到礼物的概率？最后换的概率是 $2/3$ ，一开始我直觉上以为不用换，差点就说出来不换了，还好算了下概率
- n 个人，其中一个是明星其余为普通人，明星不认识普通人，普通人都认识明星，普通人

之间是否认识未知。现给一个黑盒，可以询问一次 A 是否认识 B，问使用黑盒几次能找到明星？

面试官提示了下，最终算出了 $n-1$ 次，推导如下：A 不认识 B，那么 B 不可能是明星；A 认识 B，A 不可能是明星，每次询问必能淘汰一个人，所以最后需要 $n-1$ 次

4.3.3 矩阵运算

- 矩阵中给定两个点作为对角线组成的矩阵里的数的和（复杂度 $O(1)$ ，可简单预处理）？
- 序列【1, 0, 1, 0, 0, 1...】1 可以变成 0，0 可以变成 1 求最少多少次操作可以让 0 后没有 1？
- 矩阵，只能向右或向下，求最大路径 时间空间复杂度 (LeetCode easy 入门级的动态规划，时间复杂度 $O(M*N)$ ，空间复杂度 $O(N)$)
- 动态规划问题，一个二维矩阵，第一行一个数字，第二行两个，第 N 行 N 个，求从头到尾最大值为多少？

4.4 Leetcode&剑指 offer 原题

- Leetcode 64
- Leetcode 85
- Leetcode 104
- LeetCode 原题：铺方块（斐波拉契数列）
- Leetcode 原题：蓄水池抽样

5 编程高频问题：Python&C/C++方面

第五节

滴滴面经

编程高频问题

(整理: 江大白)

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 网络框架方面

- Tensorflow 的一些代码习惯(你是写 session.run 还是 estimator 啊，有什么区别/优缺点)?
- 你常用的框架有哪些?

5.1.2 基础知识

5.1.2.1 内存相关

- python 深浅拷贝

5.1.2.2 讲解原理

- python 传参
- python 垃圾回收，如何分代
- python 魔术方法
- python 中的反射
- python 字典底层怎么实现的

5.1.2.3 讲解应用

- 怎么使用 map 函数?
- 说说 lambda?
- 怎么得到 list 中的 index、value 值?

5.1.3 手写代码相关

- Python 实现 one-hot 特征?

5.2 C/C++方面

5.2.1 基础知识

- 重载和重写的区别？

5.2.2 手写代码相关

- 两个线程同时访问同一段代码程序？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
滴滴面经
操作系统高频问题
(整理: 江大白)
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

- awk 实现 left join

6.2 操作系统方面

6.2.1 TCP 协议相关

- 三次握手四次挥手

6.2.2 线程和进程相关

- 线程和进程的区别？

7 技术&产品&开放性问题

7.1 技术方面

- 假如有 m 和乘客 n 个司机，每个乘客有自己满意的若干司机，让你去配对司机与乘客，让

配对权值最大。这个问题从你们控制专业常用方法角度如何解决？

● 如果一个模型最后效果不好，你会从哪些方面来考虑？

● 定价策略算法岗问题：

(1) 预测一个城市不同区域一天的发单量，和滴滴业务相关，设计一些调度策略

(2) 预测一个城市不同区域的降水量（这个可以作为前一道题的特征）

(3) 判断用户感知是否顺路，给出思路（怎么获取训练数据，怎么去判断用户的感知到底顺不顺路，找一些简单的规则）

● 特征工程的题目：假如要预测一段时间内某个地方的出租车需求量，写下你认为的影响因素，然后想象怎么影响？

● 离散型属工作性和连续性属性的优缺点？连续性属性离散化的好处和坏处？（好处就是可以维度扩展，从而可以训练非线性模型，坏处就是容易过拟合，过拟合了之后怎么解决？计入正则项，或者去看离散化之后的特征，哪个特征所包含的样本太少的话就把这维特征去掉，去噪声，在特征的数量上进行制约）

7.2 产品方面

● 场景假设：预判滴滴乘客接单等待时间：需要考虑的那些因素，特殊时间天气特俗考虑

● 场景题，如何判断突发性异常事件（如追尾等）

7.3 开放性问题

● 给定北京周边各个地方的历史天气状况，来预测当天各地的天气状况？

我当时从这几方面答的：

(1) 每个地方根据历史数据+当天的天气预报相关特征+时间点（早高峰）+周几（是否周末、节假日）+PM2.5+可见度+是否下雨等等训练预测模型，我当时考虑是采用朴素贝叶斯去做

(2) 不同地方之间的天气状况会相互影响，我当时想的是根据历史数据构建不同地方的互相影响的贝叶斯网络

● 在学校建 4 个食堂 保证 4 个食堂午餐的流量基本相等 怎么选址？

19|蚂蚁金服算法岗武功秘籍

1 蚂蚁金服面经汇总资料

第一节
蚂蚁金服面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 蚂蚁金服面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 蚂蚁金服面经-35 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【蚂蚁金服芝麻信用算法实习】

(2) 全职岗位类

【NLP 算法工程师】、【计算机视觉算法工程师】、【机器学习算法工程师】、【人工智能算法工程师】

1.3 面试流程时间安排

蚂蚁金服面试流程 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	项目+延伸的基础知识
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	项目+延伸的基础知识
第三面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	简历上的项目 也会问开放性的问题
第四面	交叉面	自我介绍+项目/实习经验 +技术问答+算法编程	简历上的项目 也会问开放性的问题
第五面	HR面	基础人力问题	/
第六面	杭州终面	ppt+问答	/

PS：以上流程为大白总结归纳所得，以供参考。

其他注意点：

- 感觉实习是四面：技术+技术+技术交叉+技术&HR 混合
- 真正工作是六面：技术+技术+技术+技术交叉+HR+杭州终面
- 不过有的人也和实习那个一样，是四面，最后技术大佬+HR 是压力面
- 第一面技术面，有的人不是现场面，而是电话面
- 最后的杭州终面，一般是两个技术 leader（P9）和 一个 HR，加上两个工作人员，半小时左右。

1.4 蚂蚁金服面试心得汇总

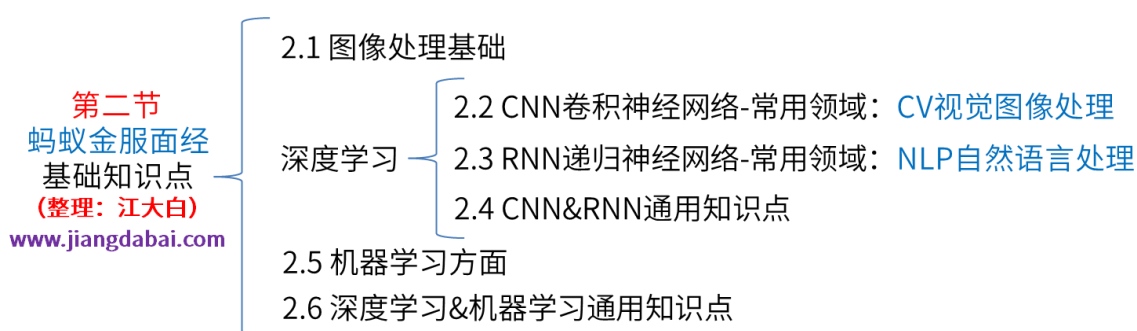
- ★ 项目的熟悉程度，其中涉及的知识点，基础知识，比如机器学习，深度学习问得很多。
- ★ 整个流程下来感觉蚂蚁非常看重的是你对于自己研究领域的思考和实践，比较注重你是否具有较高的学习热情和个人潜力

★ 蚂蚁金服面试下来的心得就是，一定要梳理好自己以往的项目，不要随意跳槽，每份工作如果不是特殊情况，尽量干满两年，简历太花可能连简历关都过不了。

★ 面试总体来说，我感觉阿里的面试包容性比较强，原意去挖掘你的优势，目前我还没遇到手撕代码，更多的是对基础的考察。

★ 很多同学交流反馈的不完全统计，阿里面试一面难度最大，之后的面试好多都是只聊简历项目。阿里的 HR 权利很大，即使面到了最后一面 HR 面还是有很大可能会被挂掉。

2 蚂蚁金服面经涉及基础知识点



2.1 图像处理基础

- Canny, 关键点是如何提取的, 关键点是如何匹配起来的?
- SIFT 原理讲一下? 如何保持尺度不变性的?

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 卷积方面

- 知道感知野吗? 什么作用?
- 卷积的作用吗?
- 你知道 depthwise-CNN 吗? 讲讲具体原理? 那 1×1 的 kernel 的作用是什么呢? 对网络 model 有什么影响?
- Dropout 怎么实现?

- Dropout 使得训练和测试输出不一致，怎么处理？

2.2.1.2 池化方面

- Pooling 的原理？
- 有哪些下采样方式？max pooling, average pooling, global average pooling。再问分别有什么用？

答：max pooling 我蠢到说提取最有特征的特征，其实就是最具有代表性的特征；average pooling 提取的是比较 general 的特征；global average pooling 用来分类的，因为后面网络的加深，full connected layer 参数太多了，不容易训练，为了快速准确得到结果，采用 global average pooling，没有参数，但是得到的分类效果跟 FC 差不多。

2.2.1.3 网络结构方面

- 为什么说 SENet 泛化性好？SENet 为什么效果好？
- 讲一下神经网络的发展？vgg, resnet, densenet 和 inception v1-4

2.2.1.4 其他方面

- BN 原理？BN 是解决什么问题的？BN 为什么有效？
- 说一下视觉的 attention？
- 梯度消失怎么造成？什么导致变得平缓，哪些函数会导致梯度消失，为什么？
- 解决 Overfitting、regulation 的方法？

(1) Dropout

介绍 dropout 的概念，问了下 train 和 test 阶段的不一样过程细节

主要讲了下 test 把 active function 的输出都乘以 P，这样就把 train 和 test 的输出期望都 scale 到了一个 range，这样才是更加准确的

(2) Batch Normalisation

BN 的原理，问什么好，好在哪里？

降低了样本之间的差异，scale 到 (0, 1)

降低了层层之间的依赖，主要体现在前一层的输出是下一层的输入，那么我们把所有数据都 scale 到了 $(0, 1)$ 的 distribution，那么降低了层层之间的依赖关系，从而使数据更加准确

- 说一下过拟合怎么造成和解决办法？ dropout 的原理，怎么抑制过拟合？ 预测阶段能不能 dropout？

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- RNN 模型, LSTM 模型讲一下？
- Transformer 与 LSTM 区别？

2.3.2 手绘网络原理

- 画出 LSTM 模型细节以及四个门计算公式？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 不平衡的 label 怎么办？

2.4.2 模型评价

- 问怎么衡量模型优劣？ 答 testing error 和 F1 Score。再问讲一下 F1 score 的计算方法？
- AUC、ROC 曲线了解吗？ 这两个曲线的应用场景是？
- AUC 曲线、precision 和 recall 分别在什么情况下使用？
- AUC 一般用于哪些地方，map 和 AUC 区别在哪里什么时候用？
- F1-SCORE 和 Auc 的区别，F1-SCORE 比较好或者 AUC 比较低是什么情况？

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

- 处理/清洗过数据吗？

2.5.1.2 特征工程

① 特征降维

- 为什么决策树之前用 PCA 会好一点？
- 具体讲一下 SVD 原理，除了降维还有什么作用？SVD 分解后怎么用于电影评分？
- 度量分解矩阵的原理？

② 特征选择

- 对于一般的问题你是怎么样处理特征的？
- 高维数据，其中有一维是时间，有缺失，如何处理？

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）

- 随机森林和 GBDT 的区别？

A. 基于 bagging：随机森林

- 随机森林的原理？
- 问 random forest 随机在哪里？
- 为什么随机森林要取 2/3 的样本构建树？
- 随机森林和 GBDT 的区别？
- 谈谈随机森林？为什么项目选择随机森林，不选择 GBDT？
- 为什么随机森林的树比 gbd 的要深一点？
- 怎么使用随机森林做特征多样性选择？

B. 基于 boosting: Adaboost、GDBT、XGBoost

- 问了解 XGBOOST 吗? 讲一下它和 GBDT 的区别?
- 说一下 GBDT 的详细过程? GDBT 如何做分类和回归?
- 从方差偏差的角度解释 bagging? 为什么随机森林泛化能力强?
- 讲一下 boosting 算法? 说一下 adaboost 是怎么更新的?
- XGBoost 或 GBDT 中怎么控制模型复杂度? XGBoost 的参数选择 (怎么调参)
- 描述一下 xgboost, xgboost 和 gbdg 的区别?

② 逻辑回归 LR

- LR 中参数的意义?
- 如何检验, p-value? LR 与线性回归比较? LR 的前提假设?
- 逻辑回归原理、手写损失函数?

③ SVM (支持向量机)

- SVM 和 Logistic Regression 对比?
- SVM 和 LR 的区别?
- SVM 适合什么样的场景? 或者说有什么限制?
- 说下 SVM, 核函数?
- 介绍一下 hinge 损失函数?

④ 朴素贝叶斯 (Naive Bayes)

- 贝叶斯公式知道吗, 什么含义?
- 朴素贝叶斯和决策树的差别, 各有什么缺点? 再加上 SVM 呢?

⑤ 决策树 (DT)

- 决策树知道吗?
- 决策树分裂节点的选择?
- 讲一下决策树的优缺点? (优点说了构造比较简单, 比较快。然后聊了 RF, GBDT, XGBoost)

2.5.1.4 无监督学习-聚类方面

- 如何确定 k-means 的 k?
- 聚类算法? k-means 的缺点?

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

- 使用过哪些损失函数?
- 损失函数怎么设计? 平方差损失优缺点? 什么时候用交叉熵和平方差区别什么时候用?
- 做损失函数的时候, 欧式距离和余弦距离的区别?
- Cross entropy loss 交叉熵损失与均方差损失 (Mean Squared Error Loss) 的差别?
- 常见的损失函数
- LR 的损失函数
- LR 支不支持非线性分类(映射到高维)

2.6.2 激活函数方面

- Relu 小于 0 的时候也会导致梯度消失怎么办? 怎么保证输入不是负数?

2.6.3 网络优化梯度下降方面

- 用什么优化方法, 梯度下降的种类, 各有什么优点?
- 神经网络为啥不用拟牛顿法而是用梯度下降?
- 说下从 sgd 到 adam 的思路?
- Adam 优化器是怎么工作的?

可以看做是 RMSProp+Momentum

具体说来, 就是它使用了一个梯度平方的滑动平均 (同 RMSProp)

然后使用的是平滑版的梯度 m , 而不是原始梯度 dx , 积攒历史梯度 (同 Momentum)。

2.6.4 正则化方面

- L1 和 L2 的区别？从贝叶斯估计的角度看，他们先验分布是什么？
- L1、L2 正则化原理，为什么 L1 产生的权重稀疏？
- 什么时候用 L1，什么时候用 L2 正则，有什么区别？

2.6.5 过拟合&欠拟合方面

- 有没有遇到过拟合？说一下什么是过拟合，解决方案有哪些？

2.6.6 其他方面

- 距离公式的局限性和改进？什么时候用，优缺点？
- 各种距离公式适用性？欧式距离，余弦距离什么时候用？

3 蚂蚁金服面经涉及项目知识点

第三节
蚂蚁金服面经
项目知识点
(整理: 江大白)
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

3.1.1.1 讲解原理

- RoI Pooling 如何操作？
- 如果有很长，很小，或者很宽的目标，应该如何处理？
- FPN 具体是怎么操作的？
- FPN 的特征融合具体是怎么做的？
- FPN 的特征融合为什么是相加操作呢？

- Soft-NMS 是如何操作的?
- Softmax 是什么，公式是什么样的?
- Softmax 的梯度是什么?
- 对 SSD 和 YOLO 有没有什么了解?
- 对 SSD 做过什么实验吗?
- 介绍一下目标检测有哪些方向，最近的一些新的进展？、
- FPN 是怎么提升小目标检出率的?
- 大目标如果有 2 个候选框和 gt 重合应该怎么处理?
- 分类和定位的不一致具体是什么?
- Faster-RCNN 比起 RCNN 有什么改进的地方?

3.1.1.2 损失函数

- Focal loss 怎么操作的?

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 与 Word2vec 区别?
- Bert 和 Elmo 在工程中存在的一些 Trick?

② Transformer

- 说一下 Transformer 的结构
- Transformer 中的前馈神经网络的构建的两种方式有什么区别（一个传统的 FFNN，一个是基于卷积的，第二种这个根本没有了解过）

③ CRF

- CRF 原理?

- CRF 怎么接在 LSTM 后面?

④ HMM 隐马尔科夫模型

- 介绍一下 HMM?

⑤ Word2vec

- 讲一下 Node2vec?
- Word2vec 的 Loss 及原理?

⑥ 其他

- 了解 seq2seq 模型? 谈谈这个模型是怎么做的? attention 是如何实现的? self-attention 了解过吗?
- 你知道 sequence to sequence 吗?

3.3 强化学习

无

3.4 机器学习方面

3.4.1 推荐系统

- FM 模型, FFM 模型, WIDE&DEEP 模型&DEEP FM 模型的区别?
- Deep FM 重点询问了细节, 画出模型结构图?
- 讲一下 FM 算法和矩阵分解的区别? 具体怎么组合? FFM 和 FM 的区别?

4 数据结构与算法分析相关知识

第四节
蚂蚁金服面经
数据结构与算法分析
(整理: 江大白)
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

- 给定一组超大数据的数组，该数组内的数据无法完全存入计算机中，问如何得到第 k 个小的数？
- 给定一个二维数组，每一行递增、每一列递增，判断目标值 `target` 是否存在？

4.1.1.2 链表

- 升序链表找中位数？

4.1.1.3 字符串

- 如果一个 `list` 里有空的字符串，怎样最快的速度把空字符串筛去？

4.1.2 树

4.1.2.1 二叉树

- 打印从根结点到叶子节点的所有路径，每条路径一行打印，最后递归方式写完？
- Z 字形打印二叉树？

4.1.2.2 堆

- N 个有序数组的排序，最后输出一个数组，采用最小堆算法解决？
- 一个文件（大小未知），每一行是一个数值，如何找出数值最大的前 100 个数？要求用大顶堆或小顶堆实现，讲实现过程？

4.1.5 排序

- 快排的时间复杂度和空间复杂度？平均多少？最差多少？
- 还有那些排序的时间复杂度是 $O(n\log n)$ ？知道排序中的稳定性吗？

- 三路快排？

4.2 算法思想实战及智力题

无

4.3 其他方面

- 不用 math 中的取平法差，判断一个自然数是不是可以开方（时间复杂度尽量低）？
- 动态规划：最长公共子序列？

4.4 Leetcode&剑指 offer 原题

- Leetcode139：单词拆分

5 编程高频问题：Python&C/C++方面

第五节
蚂蚁金服面经
编程高频问题
(整理：江大白)
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

- Python 多线程有了解吗？
- Pytorch 的卷积是如何实现的？

5.2 C/C++方面

- 写过多线程吗？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
蚂蚁金服面经
操作系统高频问题
(整理: 江大白)
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

- 多线程和多进程的区别？

7 技术&产品&开放性问题

7.1 技术方面

- 开放问题：高维数据，每个特征有标签，如何对特征进行聚类/类别划分？
- 场景题：人脸识别如果人脸图像比较少，如何用小样本学习的方法去做？
- 场景题（基于字幕文本，如何把长视频剪切成 N 个短视频）

7.2 产品方面

- 开放性问题：给你一些用户每天的相对位置信息，怎么区分他们的职业？怎么判断上线的现金贷产品的盈利能力？

20| 顺丰科技算法岗武功秘籍

1 顺丰科技面经汇总资料

第一节
顺丰面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 顺丰面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 顺丰科技面经-36 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 全职岗位类

【运筹优化工程师】、【大数据分析工程师 (用户画像方面)】、【视觉算法工程师】、【同城科技 NLP 算法工程师】

1.3 面试流程时间安排

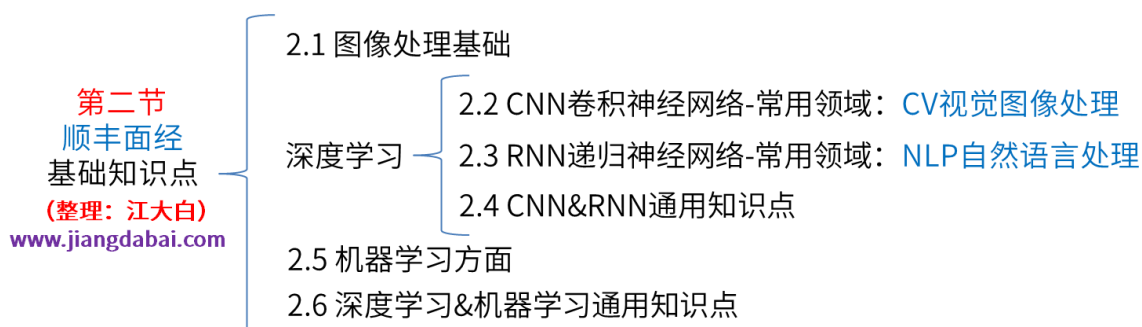
顺丰科技面试流程整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第三面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

1.4 顺丰科技面试心得汇总

- ★ 问的东西，感觉还是比较偏基础（基础还是重中之重啊）
- ★ 有些人是直接问的项目，没有问算法题或者基础题 比较关注工程落地
- ★ 感觉问的问题不是很深，都是些基础性的问题，不过待遇还都不错
- ★ 主要以简历上的项目为主，有时会加一些数据结构题

2 顺丰科技面经涉及基础知识点



2.1 图像处理基础

无

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 卷积方面

- 卷积层和池化层有什么区别？
- dropout 的原理讲一下？

2.2.1.2 网络结构方面

- mobilenet v2 的先扩张再卷积再压缩的结构？

2.2.1.3 其他方面

- 梯度消失梯度爆炸怎么解决？
- BN 的原理？

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- RNN 容易梯度消失，怎么解决？（LSTM）
- LSTM 跟 RNN 有啥区别？
- 除了 lstm，gru 了解过吗？
- LSTM GRU 了解么？
- 介绍 LSTM？

2.3.2 手绘网络原理

- LSTM 的结构？

2.4 深度学习：CNN&RNN 通用的问题

无

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

- PCA 和 LDA 讲一下？
- PCA 流程，中心化是什么，协方差矩阵怎么计算？

② 特征选择

无

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）

- 集成学习知不知道，介绍一下，说了 boosting 和 bagging？
- 做回归的机器学习方法有哪些？

A.基于 bagging：随机森林

无

B.基于 boosting：Adaboost、GDBT、XGBoost

- xgb 何时停止分裂？
- GBDT 了解吗？gbdt 与 xgb 有啥区别？
- xgb 和 lgb 的区别？xgb 和 gbdt，xgb 用的正则化？
- GBDT 树是如何生成的？
- GBDT 如何做回归？
- GBDT 的梯度是什么，对什么的梯度？

- XGB 的原理讲一下？缺点是什么？

② 逻辑回归 LR

- 线性回归和逻辑回归的损失函数知道吗，可以介绍一下吗？
- 对于逻辑回归，已经求得参数 θ ，怎么得到该参数的方差？

③ SVM（支持向量机）

- SVM 为什么是对偶问题不是直接对原问题求解？

④ 决策树（DT）

- 决策树有哪些，有啥区别？
- 决策树剪枝了解吗？简要介绍一下，打了一下 cart 的剪枝，基尼指数，问还有没有别的剪枝方法？
- 树模型怎么样防止过拟合？

2.5.1.4 无监督学习-聚类方面

- KMeans 讲讲，KMeans 有什么缺点，K 怎么确定？
- k-means 的 k 怎么确定，引申出调参的方法？
- 为什么要聚类？

2.6 深度学习&机器学习面经通用知识点

2.6.1 激活函数方面

- 各种激活函数各自优缺点？

2.6.2 网络优化梯度下降方面

- 各种优化算法比较？

2.6.3 过拟合&欠拟合方面

- 介绍一下过拟合和欠拟合？
- 防止过拟合有哪些方法？

2.6.4 其他方面

- 数据类别的分布不均衡，如何解决？

3 顺丰科技面经涉及项目知识点

第三节
顺丰面经
项目知识点
(整理: 江大白)
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

3.1.1.1 讲解原理

- fpn 的每一层的 anchor 计算公式的含义，小的特征图有大的 anchor？
- Faster RCNN 跟 RCNN 有什么区别？
- SPP，YOLO 了解吗？
- yolo v1-yolo v3 的发展与改进？
- yolo 的 backbone 为什么不用池化？
- 常用的目标检测方法讲一下？
- 三维目标检测怎么标注数据集？
- ssd fasterrcnn yolo anchor 怎么设置的有什么区别？

3.1.1.2 损失函数

- Focal loss 写一下？讲一下原理？
- IOU loss 及其改进？

3.1.1.3 手写代码

- IOU 计算

3.1.2 图像分割

- 常用的目标分割方法有哪些？
- maskrcnn roialign 和 roipooling？
- 除了 maskrcnn 还有什么可用的分割的方法？
- maskrcnn 线下采样到 14x14 然后反卷积，导致精度低，怎么解决？

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 HMM 隐马尔科夫模型

- 介绍 HMM，在实践中的应用（或工程中注意的问题）

② Word2vec

- 讲一下 word2vec？

3.3 强化学习

- GAN 网络都有哪些，loss 怎么计算？
- GAN 网络训练用了什么 trick？

3.4 机器学习方面

无

4 数据结构与算法分析相关知识点

第四节
顺丰面经
数据结构与算法分析
(整理: 江大白)
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

- 循环数组求最小值（重复元素）
- 输入一个偶数（比如 4），输出所有符合要求的数字？符合要求指的是 1 要在 3 前面，2 要在 4 前面，以此类推。方法： $(n/2, 3)$ 的二维数组记录输入，每行的最后一个数初始化为 0，表示下面要访问的是该行第 0 个元素，大思路是递归。

4.1.1.2 字符串

- 两个字符串的公共子串

4.2 算法思想实战及智力题

无

4.3 其他方面

无

4.4 Leetcode&剑指 offer 原题

- Leetcode 03
- Leetcode 原题：最大装水量

5 编程高频问题：Python&C/C++方面

第五节
顺丰面经
编程高频问题
(整理: 江大白)
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 网络框架方面

- Tensorflow 里面有没有用过 with 关键字，它的作用？

5.1.2 基础知识

5.1.2.1 线程相关

- python 的进程和线程？

5.1.2.2 区别比较

- list 和字典的区别，有序和无序

5.1.2.3 讲解原理

- python dict 是有序的还是无序的，如何实现的
- python 参数传入的是值还是引用？
- python 函数传参？
- pytorch 怎么阻断反向传播？

5.1.2.4 讲解应用

- string 怎么转成 double 格式？

5.2 C/C++方面

- 虚函数的概念，c++里析构函数的作用？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
顺丰面经
操作系统高频问题
(整理: 江大白)
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

6.2.1 常用命令

- linux 命令，取 top10IP 地址？

7 技术&产品&开放性问题

7.1 技术方面

- 顺丰业务的场景，怎么将用户写的快递单子规范化(有些快递的品名是 213453、棉衣 1 件、#46321 笔记本)
- 场景题，订单时间预估(客户下单到小哥到店取餐之间的时间预估)
- 场景题，问了预测快递件数的问题，怎么解决？

7.2 产品方面

- 现在有一个 app，里面想给用户做广告推荐，怎么做？

7.3 开放性问题

- 如果你是 leader，算法已经提前完成任务，但前端与后端没完成任务，你该怎么办？

21|依图科技算法岗武功秘籍

1 依图科技面经汇总资料

第一节
依图面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 依图面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 依图科技面经-29 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 全职岗位类

【NLP 算法工程师】、【计算机视觉算法工程师】

1.3 面试流程时间安排

依图科技面试流程 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	/
第二面	技术Leader面	自我介绍+项目/实习经验 +技术问答+算法编程	更偏向于落地
第三面	技术综合面	自我介绍+项目/实习经验 +技术问答+算法编程	/
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。

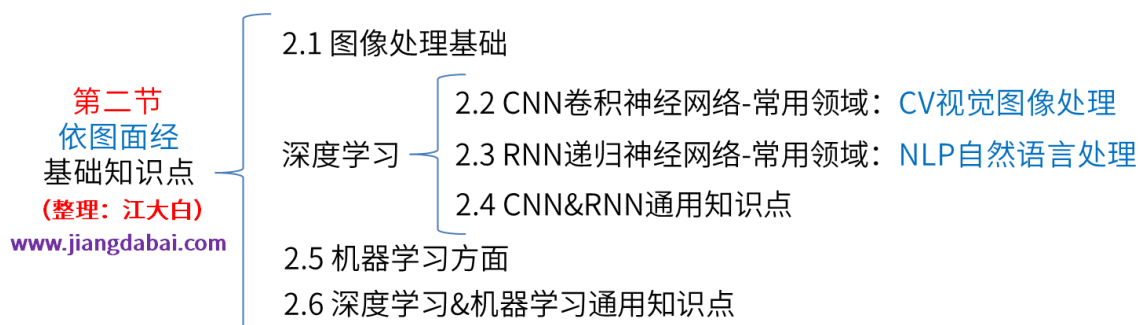
其他注意点：

- 实习岗位，有的人是 2 面+HR
- 正式岗位，有的人是 3 面+HR

1.4 依图科技面试心得汇总

- ★ 感觉对编程题和逻辑题非常注重，考察了很多这方面的题目
- ★ 也会深挖项目，使用的模型，损失函数，数据都会讲解
- ★ 手撕代码为主，不像其他公司以项目为主
- ★ 自我介绍+两道编程题+有时间问项目，三面包括最后的 leader 面都是逻辑为主
- ★ 面试官喜欢用 C++ 做算法，说第一次见到用 python 刷题的，被鄙视
- ★ 经常是稍微问下项目，然后开始做题，而且经常会根据项目问一下发散性的问题
- ★ 对依图的现场面进行一下评价，总体来说比较人性化，面试官都很友善（我做的四道题都有被提示），比较重视数学功底、coding 和推理能力，但对机器学习深度学习的基础考察的不多。

2 依图科技面经涉及基础知识点



2.1 图像处理基础

- 均值滤波的原理说一下?

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 卷积方面

- Group conv 和 separable conv 讲一下原理?
- 3D 数据直接 3D 卷积,和 2.5D 有什么区别?
- Dropout 原理?

2.2.1.2 网络结构方面

- 说一下 mobile net 和 resnet 在特征提取上的本质不同? 这两种网络使用的场景和原因?

2.3 深度学习: RNN 递归神经网络方面

2.3.1 讲解相关原理

- LSTM 和 GRU 区别?

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 模型评价

- 简历上很多的项目指标、业务场景？
- 项目中涉及各个算法的评价指标意义和数值，整体评价，数据量，怎么划分，样本不平衡？

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

- 特征工程一般怎么做？
- 针对简历上的一个数据挖掘项目，详细说明算法模型结构，一共试过哪些模型，为什么选用这个模型，数据清洗难点，特征工程怎么做的（非数值型构建三元组+TransR 生成 embedding，数值型分箱+one-hot，特征组合，构建梯度特征，归一化）

2.5.1.3 有监督学习-分类和回归方面

无

2.5.1.4 无监督学习-聚类方面

无

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

- Triplet loss, 写出数学形式?

2.6.2 正则化方面

- 正则化是什么, LN 是什么, 作用是什么?

2.6.3 过拟合&欠拟合方面

- 过拟合手段有哪些?

3 依图科技面经涉及项目知识点

第三节
依图面经
项目知识点
(整理: 江大白)
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

- MTCNN 做人脸检测的原理, 网络结构, train 的方法?
- SSD 和 Faster-rcnn 的区别?
- 问了对目标检测网络的理解以及简述 Yolo 几个版本?

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 原理?

- BERT 为什么有效，与其他模型相比呢？

② Transformer

- Transformer 的 Multi-head attention 计算细节？
- Transformer 优点？

③ Word2vec

- Word2vec、glove 和 fasttext 区别？
- word2vec 训练时出现过问题吗，比如训练后的词之间的相似性不准？

④ 其他

- 数据源如何来的，数据更新如何解决？
- embedding 方式有哪些？
- 模型预测错误的数据，为什么会错，分析过么？
- sentence pairs 模型中，为什么不直接用 score 排序？

3.3 强化学习

无

3.4 机器学习方面

- 假如用户点击了一张图片，如何根据这张图片去推荐相关视频？从数据收集，模型选取，模型评估，线上测试方面说说。

4 数据结构与算法分析相关知识点

第四节
依图面经
数据结构与算法分析
(整理: 江大白)
www.jiangedabai.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

- 旋转数组的查找？即给一个有序数组，从中间某点隔开，然后左右部分互换位置，然后问在这个数组中怎么进行二分查找。（也可能是旋转数组最小值）
- 排序数组有重复，旋转之后查找？
- 二维数组左到右上到下非递增，查找元素？注意数组维度很大时要使用二分查找思想
- 有一个字符串数组，如['A', 'B', 'C']，1) 输入两个元素，如'A', 'Z'，求这两个元素在列表里的距离，如果不存在则返回-1；2) 如果希望在 $O(1)$ 时间内查找，如何建立哈希表？
- 求数组的连续子序列，使其和最大？
- 有一个长度为 n 的数组，元素都是 $[1, n]$ 且无重复，这时随机删除一个元素，1) 求删除元素，2) 要求时间复杂度 $O(n)$ ，空间 $O(1)$ ，3) 不能改变数组，4) 如果删除两个元素，如何求？
- 一个数组，为 $1, 2, 3, \dots, n$ ，现在去掉了一个数，剩下 $n-1$ 个数。求去掉的是哪个数。注意：不准改变原数组，原数组无序空间复杂度 $O(1)$ 。第二问，如果去掉了 2 个数呢？
- 给一个数组，给一个值 s 。判断这个数组里是否存在子数组，其和等于 s ？如果有的话，有几个？ $N(n)$ 复杂度
- 两个排好序的数组，求两个数组第 k 大的数？
- 找到数组第 K 大的数（这个题有两个解法，堆排序和类似快排划分的方法，我当时只说出来一种，如果把两种说法都讲一下，并且讲一下复杂度和使用情况，肯定加分不少）
- 两个排序数组的中位数，写了 $O(n)$ 的解法，介绍了一下 $\lg n$ 的方法的思路
- 两个有序数组求第 k 大的数？

4.1.1.2 字符串

- 两字符串 a, b ，求 $a+b$ ？先填充然后再进位加，比较简单。
- 字符串数字相减？

- 给一个字符串，切割成一些回文子串，求所有可能？

4.1.2 树

- 二叉树的遍历 递归非递归
- 手写字典树

4.1.3 排序

- topK 问题，时间复杂度，代码实现？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 有 n 个人， m 个坏人。每次检查一个人无论是不是坏人都会导致这个人死亡。那么查到第一个坏人的死掉好人的期望是多少？（可以用动态规划）
- 最大水池面积，给定一个整数 n , $1..n$ 多少个唯一的 BST？
- n 个石头，每次拿 $1..m$ 个呢？编程实现？
- 1111aaaa1111ffffuuuu，输出： $[('1',4),('a',4),('1',4),('f',4),('u',4)]$ 这样的格式？

对于上面的算法题，两次操作，每次删除一个 tuple，删除 tuple 的第二个是数作为增益，删除操作可能发生合并，如删除 $('a',4)$ ，第一个和第三个 tuple 合并，求最大增益

4.2.2 智力题

- n 个石头，甲乙两人轮流取 $1..2$ 个石头，最后取石头的人输了，什么情况下甲一定赢？
- 博弈论甲乙拿石子的问题，一堆石子，甲乙每次都可以拿 $1..2$ 个，拿完最后一个的获胜，假定甲乙都足够聪明，甲先开始，什么情况甲必胜？两堆石子呢？
- 7 个石头，甲乙两个人，每次只能拿 $1..2$ 个，谁最后拿完谁赢？请问是否存在先手必胜策略？
- 平均需要抛掷多少次硬币，才会首次出现连续的两个正面？
- 2 个鸡蛋 100 层楼，从某一层开始鸡蛋扔下去会碎掉，如果鸡蛋没碎则可以继续使用。问最少扔多少次可以把那一层给求出来？

- 公司举办 Party，每个人参加 party 都会带有一个开心度，HR 希望 Party 总体开心度高，但每个人的直属领导参加这个人开心度就为 0，跨级领导不影响，多个人可能有同一个领导，HR 要确定参加 party 的人。输入两个列表，第一个为每个人的开心度，比如[2,3,4,1,1,8,5]，第二个为对应人的领导序号，如[-1,3,0,1,1,2,3]表示第一个人没有领导（-1），第二个人的领导是第 4 位，第三个人的领导是第 1 位，以此类推。求这个 party 最大的开心度为多少？
- 给定多条公交线路（可以认为是环线），以及一个初始出发点，以及目的地。求乘坐最少公交车的次数（不存在则返回-1）？
- 年利率为 4%，利滚利，多少年之后本金翻倍。就是估算 $(1.04)^n=2$ 这个等式中的 n 。在面试官提示下，又说了一种方法，大概就是， $2 \rightarrow 1.414 \rightarrow 1.44 \rightarrow 1.2 \rightarrow 1.21 \rightarrow 1.1$ ，再估计几个 1.04 联乘可以得到 1.1。

4.3 其他方面

4.3.1 数论

- 抛一枚硬币，第一次出现正面的期望？
- 抛一枚硬币，连续两次出现正面的期望？
- 证明根号 2 的无理数？
- 某大公司有这么一个规定：只要有一个员工过生日，当天所有员工全部放假一天。但在其余时候，所有员工都没有假期，必须正常上班。这个公司需要雇用多少员工，才能让公司一年内所有员工的总工作时间期望值最大？
- 最大子区间，随机数截断，并求出均值？
- 和为某一值的连续子串的个数。我说用累加法再求差，面试官说数组都是正数，可以不用额外空间吗？
- 大数加法，包含小数和正负号？
- 输入一个列表，删除相邻的相同元素，比如输入[1,2,2,3]，返回[1,3]，输入[1,2,2,2,3]，返回[1,2,3]，输入[1,2,4,4,2]，返回[1]？

4.3.2 概率分析

● 一枚硬币，抛 n 次，问正面数大于反面数的概率？ n 是常数，不是无穷大。这题分奇数偶数讨论。偶数时写出概率的式子。奇数答案是 $1/2$

● 一副扑克牌 54 张，给 3 个人发牌，把大小王发到了同一个人的概率？

● 4 只鸭子，随机在池塘中行走，请问四只鸭子能分到同一个半圆的概率是多少？

● n 个骰子扔一下，点数加起来等于 m 的概率？

● 求圆上任意三个点组成一个三角形是锐角三角形的概率？

● 抛 n 次硬币，求反面比正面少的概率

● 掷三个骰子，掷出来的三个值之和为 10 的概率。我用的笨方法，列举出所有情况算概率？

掷四个骰子，掷出来的四值之和概率最大的是哪一个。猜的 14，猜对了。

第二问中的掷出来是 14 的概率有多大？没有任何思路，后来面试官告诉我可以用动态规划来做，在面试官的提示下，写了动态转移方程。

4.3.3 矩阵运算

● 矩阵 $m \times n$ ，从左上角到右下角，路径和最大？

● $n \times m$ 矩阵，填充成 S 型走位的正整数？

● 01 矩阵，滑动窗口翻转修改为全 0 矩阵？

● 已知矩阵高 h ，宽 w ，蛇形矩阵？写代码把他画出来。 第二问，给定矩阵内一点 y, x 在 $O(1)$ 时间内给出该点的值？

● 给定一个二维矩阵，计算其子矩形范围内元素的总和？该子矩阵的左上角为 $(row1, col1)$ ，右下角为 $(row2, col2)$ 。 要求求和复杂度为 $O(1)$ 。

● 给定一个矩阵，里面元素可以为负数，0 或者正数，求最大子矩阵的和，子矩阵可以为一个数字，或者方阵矩阵。

● 给你一个矩阵，里面只有 0 或者 1，假设给定其中的一个位置，它的相邻 8 个位置有 1，则认为他们是连通的，求最大连通区域 1 的个数。

● 一个矩阵，从左到右，从上到下是递增的，可能会有相等的，给一个目标值，请查询其在矩阵

中的位置，如果不存在则输出 $(-1, -1)$ ，如果有多个，则随便输出一个，请使用小于 $O(m+n)$ 的解法。

- 给定一个 $n*m$ 的矩阵，从中找出 $a*a$ 的矩阵，使得和最大。返回最大的数值。
- 计算一个 $n*m$ 矩阵，每个是左上角元素和的值。然后再遍历计算。
- 给定一个 $M*N$ 的矩阵，求一个子矩阵，使子矩阵的和最大？
- 输入一个二维矩阵，寻找一条从左上角到右下角的路径使得经过的路径值加起来最大？（只能向右和向下走）

4.4 Leetcode&剑指 offer 原题

- LeetCode 33
- LeetCode 52: N 皇后问题
- Leetcode 80: 二维数组全是 0 或 1，求全为 1 的最大子矩形面积
- Leetcode hard: 矩阵中求全 1 的最大矩形的面积

5 编程高频问题：Python&C/C++方面

第五节
依图面经
编程高频问题
(整理: 江大白)
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

- 为什么用 python 的 multiprocessing，为啥不用线程？

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
依图面经
操作系统高频问题
(整理: 江大白)
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

- 线程进程区别？

7 技术&产品&开放性问题

7.1 技术方面

- 关于深度学习/机器学习的；通常情况下标签都是无关的，比如车辆图像的类别这种，假如标签是有一定关系的，比如给定一张图象的标签是这个人的年龄，可以分为老人、中年人、小孩等，他们是有次序关系的，如何利用这些信息？

答：（仅供参考，不一定对，欢迎评论提出意见建议）可以用向量的余弦构建 loss。

22|旷视科技算法岗武功秘籍

1 旷视科技面经汇总资料

第一节
旷视面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 旷视面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 旷视科技面经-36 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

- (1) 实习岗位类

【图像算法实习生】

- (2) 全职岗位类

【图像算法工程师】、【算法研究员】

1.3 面试流程时间安排

旷视科技面试流程 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	项目细节问的很细
第二面	技术综合面	自我介绍+项目/实习经验 +技术问答+算法编程	项目细节问的很细
第三面	技术Leader面	自我介绍+项目经验+公司发展	项目细节及基础 也会问项目中的优点和难点
第四面	HR面	基础人力问题	/
第五面	终面	自我介绍+项目经验+公司发展	/

PS：以上流程为大白总结归纳所得，以供参考。

其他注意点：

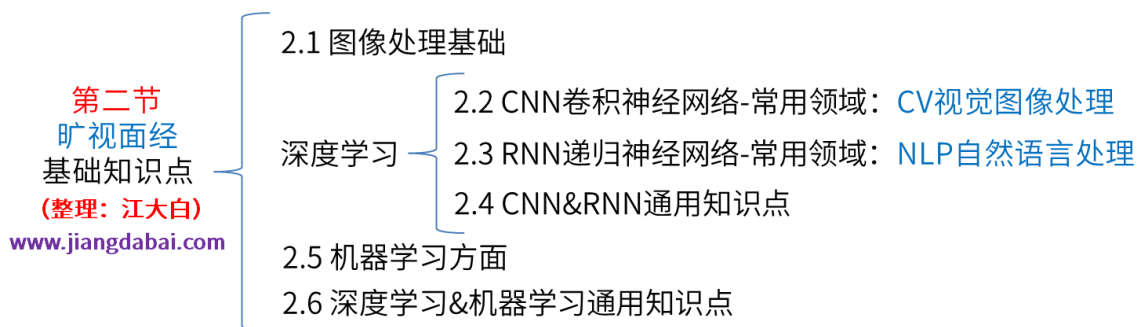
- 最后的终面可能是某个大佬，比如孙剑，喜欢问细节，问得特别细，对细节比较在意。开始仍然是介绍研究经历，不过和其他所有面试相比，最大不同在于孙剑老师会不停的发问，不断 challenge。之后便是向孙剑老师提问。整个下来大概 45 分钟。
- 最后的终面，也可能会让你基于项目在白板上画模型结构讲解（个人感觉比较侧重考察交流能力）。也会问开放性题目，自己研究领域的一些问题，还有自己的看法以及解决思路。

1.4 旷视科技面试心得汇总

- ★ 一面问得很细，每一个项目的每个细节都要问为什么不用其他方法？
- ★ 问得很深入，基本都要非常理解才行，提到某个细节之后可能会深入问这个细节更细节的东西，千万别在回答的时候给自己挖坑。
- ★ 一面问项目技术问得很多，二面二面考基本功，数学算法和 ML 的熟练运用能力，问算法逻辑设计问得很多。

- ★ 三面问整体的项目，前沿知识，比如（最新的网络结构、精度最高的目标检测模型等）
- ★ 三面一般会问一些本质，感觉和一二面问技术细节不同，你可以不知道细节，但一定要清楚为什么这么做。同样会一些研究前沿的问题，主要还是要有自己的思考，调参侠真的不可取。懂一点高级的抽象数学还是最好。
- ★ 每一面都是概率/代码题，加项目实习经验，另外可以准备一下：个人职业规划、优缺点等以及一些提问问题（技术层面和其他层面都可以准备一下）。
- ★ 会结合自己负责的一个项目，在该项目中是如何安排整个进度以及项目开展中比较难的部分。
- ★ 总的来说，就是基础 coding 能力必须过关，至少 offer 上的题目以及 leetcode 一些关于数组和字符串的中等题得会，问科研部分会问的很深，很细节，会让在纸上画出 model structure 并给出原因，另外就是，最好能在自己领域做到相对深的情况下，去开阔视野，多 follow 其他领域的 paper，尽量又深又广。做题过程中面试官会一直引导，就算刚开始不会也不要慌，重点是考察思考问题的方法和角度。

2 旷视科技面经涉及基础知识点



2.1 图像处理基础

- 图像旋转、旋转矩阵、像素点怎么填充？
- SIFT 算子的尺度空间因子怎么设置？
- 讲讲 Canny 算子的算法？

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 卷积方面

- 空洞卷积的具体实现？
- 两个卷积核大小为 (3,3)，步长为 (2, 2) 的卷积层堆叠，上层感受野大小是？
- Deformconv 的原理？
- 问 mobilenet_v2 1*1 卷积作用？
- 常用扩大感知野的方法

2.2.1.2 网络结构方面

- Inception v1-v4 的区别、改进？
- VGG16 和 Resnet152 哪个参数量多？
- 说说 Densenet 的优势？
- 谈谈 ResNet 结构？ResNet 的特点
- 讲一下 SENet 的模块，为什么有性能提升，有什么好处？
- 模型轻量化了解吗，讲一下？mobileNet v1, v2。
- 介绍了解的神经网络(说了 lenet, alexnet, googlenet, inception v1-v4, resnet, densenet, mobilenet)
- shufflenet 结构

2.2.1.3 其他方面

- 讲讲梯度消失问题及其应对方案？(BN、Relu、初始化)
- BN 优点，原理？各参数的意义？NCHW 的特征图 BN 层参数？
- 什么是 BN？BN 有啥优点？BN 层怎么计算？讲讲 BN 的细节（过程，公式，作用）？为什么 BN 可以加快优化算法的速度？

- 问 BN，详细介绍，问是否可以改进，训练和测试过程是否一样，测试中需要注意什么。均值和方差的计算是无偏的还是偏的？
- Data augmentation 怎么处理的？Noise Label 怎么做？
- 反向传播的时候为什么会梯度消失？
- 多标签分类怎么解决，从损失函数角度考虑？
- 零样本分类问题。如果测试时出现一个图片是训练时没有的类别，怎么做？
- 一张图片多个类别怎么设计损失函数，多标签分类问题？
- 图像分类模型 ResNet101 相比 ResNet50，residual block 添加到哪里？为什么要这样添加？
- 细粒度分类：bilinear CNN，细粒度分类 主要的难点及目前的方法，其他的细粒度 task 了解吗？

2.2.2 数学计算

- $n \times n$ 的 feature map 上执行 $m \times m$ 的最大池化，步长 1，padding $m/2$ ，设计算法并求时间复杂度

2.2.3 手写算法代码

- 手写 BN 的实现。注意 BN 的 mean 和 std 是在哪个维度求梯度的，mean 和 std 是滑动平均的值。基于 numpy 实现。
- 用你熟悉的框架写一下 Unet 结构
- numpy 完成 avepooling 计算（积分图）

2.2.4 激活函数类

- 问了 softmax 损失函数

2.3 深度学习：RNN 递归神经网络方面

2.3.1 讲解相关原理

- RNN 怎么反向传播？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 输出值变成 Nan 了怎么办？估计是超过 float32 的数据范围了，需要对中间的数据进行归一化，比如 BN，sigmoid 函数等等。
- 对 attention 有系统了解过吗？讲一下各个 modelde 异同。SEnet, SKnet, CBAM, Non-local。

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

- PCA了解不,其优化目标是什么？最大化投影后方差+最小化到超平面距离--机器学习方面。
- PCA 具体过程？PCA 白化是什么？为何选择方差最大的方向进行新的特征选择？
- 介绍 PCA，特征值分解，奇异值分解等等（描述了这两个分解的意义，特征值和特征向量的用途以及这两个分解的应用）。

② 特征选择

无

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）

- 讲讲 adaboost 和 random forest 的相同之处和不同，各自应用范围，实际应用选择？

② SVM（支持向量机）

- 对 SVM 的理解，为什么要用对偶问题（二次规划+核化）具体讲一下为什么要核化，核化的过程？
- SVM 怎么用于回归？
- SVM 的损失函数，特点，对偶问题求解？

③ 决策树（DT）

- 决策树怎么做回归？没答出来。后来在他的引导下说是给回归定义一个损失函数，比如 L2 损失，可以把分叉结果量化；最终的输出值，是分支下的样本均值。

2.5.1.4 无监督学习-聚类方面

- Kmeans 的原理？

2.5.2 手推算法及代码

- 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

- 详细讲一下 BCE LOSS, Dice Loss, focal loss, triplet loss, center loss。
- 写 Triplet Loss 代码？
- 交叉熵公式写一下？

2.6.2 网络优化梯度下降方面

- 梯度下降法和牛顿法区别？
- Adam 和 SGD 区别，RMSProp 优化算法？
- SGD 的形式

2.6.3 过拟合&欠拟合方面

- 过拟合是怎么造成的，如何解决？

3 旷视科技面经涉及项目知识点

第三节
旷视面经
项目知识点
(整理: 江大白)
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

3.1.1.1 讲解原理

- 讲一下 FPN 为什么能提升小目标的准确率？
- 说一说 SSD 具体是怎么操作的？
- FPN 多尺度层中的 box 如何进行 ROIpooling？
- 讲一下 DL 中目标检测的类别和特点？（one stage、two stage）
- 为什么 two stage 比 one stage 慢，为什么 two stage 比 one stage 精度高？one stage 在哪些具体方面检测精度不高（ROI+default box 的深层理解）
- One-stage 检测针对负样本太多有什么解决办法？
- 目标检测算法中 one-stage 方法中为了解决正负样本不均衡，除了 focal loss 和 xxx 还有

什么方法？

- 目标检测 one-stage 方法中 anchor 预测出的位置是有偏移量的，而类别置信度是针对原始位置的特征并不是偏移后的特征，提出解决办法。
- Faster RCNN 和 SSD 区别？
- Anchor 的正负样本比是多少？
- 为什么项目中用 Faster R-CNN+FPN, Faster 和 YOLO 对比；为什么叫单步法，两步法？
- Faster RCNN 中 RPN 改成 SSD 如何改进，如何选择特征层做 ROI？
- 什么场景适合检测，什么场景适合分割？
- anchorfree 和 anchorbase 说一下，你说到 fcos 看过源码吗，说一下细节？
- roi pool 和 roi align 区别？

3.1.1.2 损失函数

- 讲一下 Focal loss 的 2 个参数有什么作用？
- yolo 系列介绍一下，重点说一下损失函数？
- 除了 smoothl1 bbox 的其它损失函数有什么了解吗？

3.1.1.3 手写代码

- IOU 和 NMS 怎么算，写代码
- 给两个矩形，算 IOU
- 给一张二值图像，1 的四邻域有 1 称作联通的白块，找出所有联通白块的外界矩形

1 1 0

1 0 0

0 1 0

输出[(0,0,1,1), (2,1,2,1)]

3.1.2 图像分割

- 说一下 deeplab? 它与其他 state of art 的模型对比, 它的亮点, deeplab 还可以做哪些改进? 介绍 deeplabv3, 画出 backbone, deeplabv3 的损失函数。
- 串联与并联的 ASPP 都需画出。论文中认为这两种方式哪种更好?
- mIOU(图像分割的通用评估指标)的代码实现? 使用 numpy (我直接用了 python)
- 常用分割网络

3.1.3 异常行为识别

- 了解其他异常行为检测的算法吗?
- 异常行为定义, 实际部署环境, 怎么检测, 速度如何保证?

3.1.4 姿态估计

- 姿态估计算法中 Top-down 方法与 One-stage 方法的不同, top-down 方法准确率高的原因是什么? (比较侧重挖掘对该领域研究的理解深度。。)

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

3.2.1.1 讲解原理

- CRF 后处理的目的?

3.3 强化学习

无

3.4 机器学习方面

无

4 数据结构与算法分析相关知识点

第四节
旷视面经
数据结构与算法分析
(整理: 江大白)
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

● 一个数组里有很多数，只有一个只出现了一次，其他都只出现了两次，怎么找？那如果要找两个只出现了一次的数呢？

● 一个正整数组成的数组，分成连续的 M 段，问每段的数之和的最大值最小是多少？

例如：a=[10,6,2,7,3],M=2,答案为 16，两段分为[10,6][2,7,3]

(x0,x1),(x1,x2),(x2,x3),(x3,x4),...,(xn-1,xn)矩阵计算的时间复杂度，有什么优化方法

● 求一个数组中只包含 0,1 使得其中 0,1 个数相等的最大子数组？

● 有一个数 target，一个数组，数组中两个数的和是这个 target，请找到这两个数？

● 连续子数组的最大和？

● 输入一个二维数组，输出是这个二维数组的最大子数组(也是二维)。最大指的是元素累加和最大。如果是本身，那么返回这个二维数组。

● 给一个数组，给一个值 s。判断这个数组里是否存在子数组，其和等于 s？如果有的话，有几个？ N(n)复杂度

● 两个有序数组，返回并集的中位数？

● 一个数组，为 1, 2, 3, ..., n，现在去掉了一个数，剩下 n-1 个数。求去掉的是哪个数。注意：不准改变原数组，原数组无序空间复杂度 O(1)。第二问，如果去掉了 2 个数呢？

- 数组 shuffle 以及证明算法 shuffle 后产生的数组排列是均匀随机的？
- 输入一个二维数组，和任意两点坐标，输出这两点坐标(左上点，右下点)围成的矩形内所有元素和，用积分图方法。
- 对于给定的数组 A，每次可以取出任意一个数 $A[i]$ ，同时要删除数组内所有等于 $A[i]-1$ 和 $A[i]+1$ 的数，直到数组为空，可以取出所有数的最大和为多少？
- 用 numpy 实现 $m \times 2$ 的矩阵中每一组数 (x_i, y_i) 与 $n \times 2$ 的矩阵的每一组 (x_j, y_j) 的欧氏距离之和，即有 $m \times n$ 组欧氏距离，其中 i 从 0 到 $m-1$, j 从 0 到 $n-1$

4.1.1.2 链表

- 链表反转？
- 链表的倒数第 k 个结点（双指针）
- 判断两个链表是否相交？
- 求两个链表的第一个公共节点？
- 链表排序（要求快排）
- 链表归并快排？

4.1.1.3 字符串

- 一段字符串的句子，由多个单词组成，返回颠倒后的句子？（单词不颠倒）
- 找到最长回文子串？
- 两个字符串序列的最长公共子序列？
- 寻找目标字符串的子串（可不排序）

4.1.2 树

- 堆介绍，插入元素时调整的时间复杂度（变成二叉树，递归定义）堆排序、其他排序方法介绍和特点（按时间复杂度分了三样去介绍），最常用哪种？

4.1.3 排序

- 实现归并排序
- 编程求数组中的 Top K 大的数？
- 求一个数列中两个元素的最大和，找到这个两个元素？（Top K 问题）
- 手撕 topk，用的堆

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 求一个数的方根，精度 0.01？二分加递归
- a 的 b 次方，最优解法
- 一个数开平方，如果开不了就取最近的平方根值，比如 10 开平方是 3
- 十进制转二进制
- 区间合并就是[1 2 3 6 10 11 12]合并为[1-3, 6, 10-12]
- 连续子序列的最大和？

4.2.2 智力题

- 房子 500 万，每年涨 10%，程序员工资 100 万，不涨，问多少年能全款买房（几秒钟估算了一下，永远买不起…）
- 一个人从原点出发，可以往左走可以往右走，每次走的步数递增 1，问能不能到达一个位置 x？如果能，给出走的步数最少的方案？
- 10 只小白鼠，986 瓶水，其中一瓶是毒药，如何一次性找到哪瓶是毒药？
- 掰巧克力， $m \times n$ 格，掰成一小块一小块，不能叠着掰，最少需要掰多少次？（答案： $m \times n - 1$ ）

4.3 其他方面

4.3.1 数论

- 用拉格朗日乘子法将约束优化转化为无约束优化，直观解释一下拉格朗日乘子法？
- 设随机变量 $X_1, X_2, \dots, X_n$ 相互独立，且都服从 $(0, \theta)$ 上的均匀分布。求 $U = \max\{X_1, X_2, \dots, X_n\}$ 数学期望？
- 二维直角坐标系上，给定 N 个点的坐标（float 型），一个点 C 的坐标（float 型），一个整数 M 。问：找一个正方形，它以 C 点为中心，且四条边分别与 x 轴和 y 轴平行，其内部至少包含 N 个点中的 M 个点（在边上也算），问这样的正方形边长最小是多少？
- 给定一个 $[0, 1]$ 的均匀分布，求圆周率？
- 对极几何公式推导？
- 一个是两个均匀分布，求二者差的绝对值的期望？
- 已知 a, b 都服从均值为 0，方差为 1 的正态分布，求 $\max(a, b)$ 的期望。
- 一个点位于零点位置，可以向负方向移动，也可以向正方向移动，概率都是 $1/2$ ， X_t 表示移动 t 步以后点 p 到零点的距离，计算 X_t 的期望， X_t^2 的期望？

4.3.2 计算几何

- 怎么求一个三角形外接圆，三条边垂直平分线的交点即圆心，圆心到顶点的距离为半径？

4.3.3 概率分析

- 抛一个不均匀的硬币，设计策略能得到 $1/2$ 的概率（抛两次）如果要求得到 $1/3$ 和 $2/3$ 呢？设计策略（抛四次，我想着抛 6 次，小哥哥提醒了）
- 一枚硬币，抛 n 次，问正面数大于反面数的概率？ n 是常数，不是无穷大。这题分奇数偶数讨论。偶数时写出概率的式子。奇数答案是 $1/2$
- 一副扑克牌 54 张，给 3 个人发牌，把大小王发到了同一个人的概率？
- 一个半径为 R 的圆，用多少个 $R/2$ 的圆可以铺满？
- 随机选 3 个点，使他们的平方和为 1，要求选取的点等概率？

- 如何判断一个点在一个凸多边形里？
- 三个点在同一半圆上概率？
- 从 n 个数字中均匀采样出 k 个数字，无重复，并证明均匀性。如果 n 未知如何实现？如果概率都为 k/n 如何实现？

4.3.4 矩阵运算

- 矩阵 $m \times n$ ，从左上角到右下角，路径和最大？
- 已知矩阵高 h ，宽 w ，蛇形矩阵。写代码把他画出来？第二问，给定矩阵内一点 y, x 在 $O(1)$ 时间内给出该点的值？
- 最大子矩阵和

4.4 Leetcode&剑指 offer 原题

- Leetcode215: 数组中第 K 个最大元素& 栈的最小值

5 编程高频问题：Python&C/C++方面

第五节
旷视面经
编程高频问题
(整理: 江大白)
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 网络框架方面

5.1.1.1 Pytorch 相关

- Tensorflow 与 Pytorch 的区别？

5.1.1.2 Tensorflow 相关

- TensorFlow 怎么在网络结构实现一个 if 判断？

- 用 tensorflow 计算 lou，写的 python 版本？

5.1.1.3 其他

- 用什么框架，各个框架的实现机制有什么不同？

5.1.2 基础知识

- Python 多线程多进程？
- Python 除了__init__ () 函数，还有什么自带参数的类的初始化？

5.1.3 手写代码相关

- 用 Python 实现 $n \times k$ 和 $m \times k$ 的欧氏距离，不能用循环？

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 区别比较

- Cuda 里面 host 和 device 有啥区别？

5.2.1.2 讲解原理

- C++一些语法特性平时项目中会用到吗，比如多态、继承等？
- 什么是动态联编？
- C++，问了 map 等是用什么实现的。答：红黑树。（面试官：好了，我也不问你红黑树了），那你在想想还有其他实现的方法吗？平衡二叉树，差不多说了一下。
- 构造函数和析构函数的作用，析构函数何时被调用？
- C++中 sort 底层是什么，是否稳定，还有哪些稳定排序算法？

5.2.1.3 讲解应用

- Set map 低层实现的数据结构是什么？

5.2.2 手写代码相关

- 用 C++ 将一个类改造成线程安全的类？
- map, m[1]="str";这句底层是怎么实现的，发生了什么？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
旷视面经
操作系统高频问题
(整理: 江大白)
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

- 问堆和栈，说了说栈的特点，怎么用的。
- 在计算机系统中，栈有哪些用处，具体解释了一下？我说了线程和进程。

6.2.3 常用命令

- Linux 的一些常用命令：我说了几个。他又问怎么按时间顺序打印出文件列表，按文件大小打印文件列表？

7 技术&产品&开放性问题

7.1 技术方面

- 各种深度学习训练技巧以及涨点的经验，问对此有什么 insight，如何更好的解释？
- 无监督学习了解哪些？
- 1000 个视频，无标签，怎么分类预测？

- 多标签识别怎么做?
 - 数据不均衡怎么处理?
 - 项目改进点，数据量少如何解决?
 - 在原数据集训练权重上 finetuning 效果如何?
 - 特征检索加速方式、模型加速方式?
 - 场景题：许多图片数据，这些图片是由很多网络分割的，就像棋盘一样。每一个格子中可能存在一条小斜线（因为是直线，所以实际由两个端点就可以确定）。要求设计一个网络来检测出这张图片中的这些小短线？
- 要求：自己定义图片的尺寸，网络的模型，loss，评价指标。问的比较细，每一步的实现细节。
- 开放问题：一个二分类，如果只有正样本怎么做？

23|微软算法岗武功秘籍

1 微软面经汇总资料

第一节
微软面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 微软面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 微软面经-34 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【机器学习实习生】、【微软苏州 SWE 实习生】、【NLP 实习生】、

(2) 全职岗位类

【Bing 团队算法工程师】、【苏州算法工程师】、【STCA 算法工程师】、【苏州微软 Multimedia Search 组算法工程师】

1.3 面试流程时间安排

微软面试流程-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	/
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/实习经验 +技术问答+算法编程	主要是对于项目的深度思考
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：

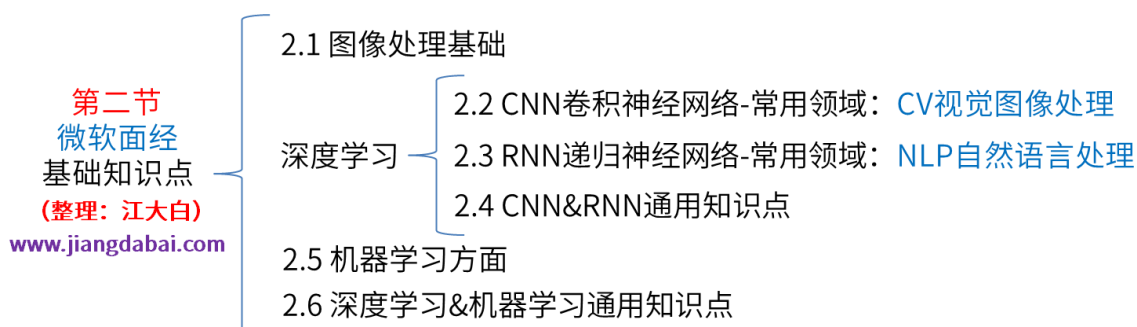
- 实习时，是笔试+3轮远程面试，笔试是4道题，远程视频面，一共三面，每次50分钟，都是先讲15分钟左右项目，然后用skype共享桌面做题。
- 有的面试最后还有Boss面
- 微软的社招面试通常是先进行一轮电话面试，面试通过的话才会邀请进行现场面试
- 电话面试之后会约现场面试，通常会安排5-6轮的面试，每轮一小时，前3轮是基础面，面试结束后面试官商量决定要不要进行后续的面试，当然如果表现比较差，也可能在某一轮直接结束。

1.4 微软面试心得汇总

- ★ 想去微软的话，编程功底比专业知识重要的多，基本leetcode easy, mid都要刷一刷
- ★ onsite面试都是在黑板上写代码，写完告诉面试官逻辑就行。三面在collabedit上做，用自己的IDE调试，需要共享屏幕。
- ★ 做完每道题都会问你怎么测试自己的代码是正确的，考虑哪些边际情况。

- ★ 算法之前一定要冷静思考一下，想一想可能有的坑，要多与面试官沟通
- ★ 有的时候，那一轮面试遇到了英国面试官就变成了英文面试，所以英文介绍最好也要准备一下
- ★ 微软更注重的是编程能力，想面微软的同学建议好好刷题，微软一般每一面都有算法题。
- ★ 面试重点还是在做题上，而且大多都是剑指 offer 和 LeetCode 的原题，所以感觉研究面经不如多刷几个题。另外最好提前准备一个英文自我介绍和项目介绍，虽然我没遇到，但很多人遇到了。
- ★ 微软的面试整体偏向基础，英语能力考察仅限于个人简介和项目描述，如果运气好的话都是中国的面试官，没有英文面试。
- ★ 投递简历之后会有 hr 先和你聊一轮，要求做一个一分钟的英文自我介绍，然后会对英文能力做一个整体评估，告诉你应该怎么准备可能的英文面试。

2 微软面经涉及基础知识点



2.1 图像处理基础

无

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 卷积方面

- 反卷积具体怎么实现的？

- 为什么 dropout 能减少过拟合?

2.2.1.2 其他方面

- 问 backpropagation 的基本公式，问每一层之间是否能独立传播?

2.3 深度学习：RNN 递归神经网络方面

- LSTM 模型结构? GRU 知道吗? LSTM 参数量是多少?

2.4 深度学习：CNN&RNN 通用的问题

- 什么是过拟合吗? 如何解决过拟合? 如何判断过拟合了?

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

无

2.5.1.3 有监督学习-分类和回归方面

- 介绍 bagging 和 boosting?
- GBDT 原理说一下?
- xgb 和 gbd 的区别 (几乎必问的题目，提前准备一下，说的要有条理，有哪些算法优化，哪些工程实现优化，可以适当扩展提一下 lgb)
- 深度学习网络 Factorize Machine 相对于线性模型有什么好处? Spark 用过吗?
- 如何构建一个分布式机器学习框架?

2.5.1.4 无监督学习-聚类方面

无

2.6 深度学习&机器学习面经通用知识点

无

3 微软面经涉及项目知识点

第三节
微软面经
项目知识点
(整理: 江大白)
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

- RCNN, Fast RCNN, Faster RCNN, Yolo, 问了我具体的 yolo 的那个 anchor, 反正好多具体的东西?

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 有几种 Embedding 编码, 分词方法?

② Transformer

- Transformer 结构讲一下?

③ Word2vec

- Word2vec 中, 负采样相比层次化 softmax, 有什么优缺点? 层次化 softmax 能保证概率

归一化吗？

- word2vec 原理？

④ 其他

- 文本分类的方法有哪些，深度学习和非深度学习的方法都说一下？
- 文本相似度计算方法有什么，当我说完后，面试官说你说的基本都是深度学习方面的，经典的 NLP 方法知道有哪些吗？
- Fasttext 和 textCNN 说一下原理？
- CBOW 和 skip-gram？

3.3 强化学习

无

3.4 机器学习方面

3.4.1 推荐系统

- FM 算法、ALS 矩阵分解、协同过滤算法都说一下，并说下优缺点？

4 数据结构与算法分析相关知识点

第四节
微软面经
数据结构与算法分析
(整理: 江大白)
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

- rotate 一次的数组，找 target，例如 [3,4,0,1,2] 找 4 所在的位置，如果不存在返回-1，要求 $\log n$ 时间 (LeetCode medium 原题，直接二分即可)
- 给一个有重复数字的有序数组和一个数 x ，找出 x 在数组中最左和最右的下标，不存在的返回 [-1, -1]?
- 在排序数组中用二分查找找到某数字的第一个位置?
- 无序数组找第 k 大的数
- 二维数组找递增 target
- 有一个数组元素[a0, a1 ...]，从数组中找出连续的数组和为最大。
- 有一个常数 n ，有一个数组元素[a0, a1 ...]无重复元素。从数组里面找出所有可能的组合加和是 n ，并且输出。
- 有一个数组（对，全是数组题目），从数组中找出连续数组乘积最大
- 连续子数组的最大和?
- 假设一个数组只有"a"和“b”两种 string 组成。如何重新安排数组，使得最多有 3 个 a 相邻，3 个 b 相邻。如果不能安排，返回 None
- 求数组最大值时，从前往后遍历，候选值会被更新若干次，求这个次数的数学期望（说思路，我算出来是 $1+1/2+\dots+1/n$ ）
- 两个有序数组找第 k 大?

4.1.1.2 链表

- BST 转双向链表?
- 单链表找交点?

- 如何判断两个链表是不是有交点？
- 实现两个链表排序？
- 链表栈哈希表的区别？哈希表的原理解决冲突什么的，排序函数以及分别适用的场合？

4.1.1.3 栈

- 给定一个温度的时间序列，判断高于当前温度的那一天在几天后出现？（先写了个 n^2 的，面试官说复杂度太高，短路想不出来，他说用栈，改了个 $O(n)$ 的）
- 用栈模拟队列？
- 最小栈，空间优化？

4.1.1.4 字符串

- 两个字符串的编辑距离（增加、删除、替换一个字符距离为 1） [动态规划]
- 中文字符串，比如一千五百亿八千九百万六百二十这种形式，转换成 long long 的整数？要我考虑很多非法输入。
- 找出字符串中所有连续字母的 subset？
- 字符串中最长连续不重复子序列长度？ [一次遍历，额外数组记录字符最后一次出现的位置]
- 一个字符串 切分成多个回文串，返回所有可能，如 aab 要返回 $[[aa,b],[a,a,b]]$ 3
- 给定一个字符串，判断是不是合法 IP 地址？
- 假设有 a, b 两个 int，转成二进制后 $c = a | b$ 。假设从 0→1, 1→0 理解为一个 action。最少需要多少个 action 计算 $c = a | b$ ？
- 给定一串字符串，输出该串字符串的全排列（完全相同的字符串算一个），同时需要满足条件『相邻字符不能相同』。

4.1.2 树

- 非递归前序遍历二叉树？
- 前序遍历、中序遍历、后序遍历，知道那些可以恢复二叉树？原因？

- sorted linked list 从头到尾翻转一次？二叉树搜索。
- 字典树

4.1.3 排序

- 写归并排序，非递归
- Top K
- 写一个快排，非递归？
- 将一堆大小写字母根据大小写排序？
- 写快排，问时间复杂度和空间复杂度，然后要求输出排序后对应元素的原下标？
- 堆排序以及很多变体？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 丑陋数变体。手撕代码，推出了 int 型的丑陋数上限并分析复杂度？
- 给一个 100 万规模的词典，一个长文档，如何快速从里面标注出所有的词，写一下代码？
- 给一句英文，in-place 把每个词顺序翻转；给一个全排列的某个情况，找到这个 permutation 在全排列中的 index？
- 给定一个字符串，判断是否是有效的 IP 地址？（提示，有效的 IP 地址格式是 xxx.xxx.xxx.xxx, xxx 在 0-255 之间）（输入不保证都是这种格式，要自己判断，同时 001 这种是否有效要问面试官）
- 找最深的左括号，follow up：括号匹配。
- 【二极管能显示的数字】

一组二极管有七个，能表示 0-9 十个数，现在有 n 组二极管，每组二极管中有一些亮，有一些灭，灭可能是因为坏了也可能是因为不需要亮。要求出这 n 组二极管能显示的所有数字的组合

输入：二维数组，每行代表一组二极管，每行七个，0 代表灭，1 代表亮

输出：所有可以显示的数字组合

例：现在两组二极管，显示的是 23，那输出 23，83，88，28，用了个回溯

- 从(0,0)出发，每一步有上下左右和停留五种选择，问 k 步之后回到原点的概率有多大，不能超出边界？（先用了回溯，后面用了 dp，面试官说我最开始不该暴力，应该直接给出最优解）
- 两个长度为 m 的无序数组 A, B，对于任意不相交的区间 ab 和 cd， $val[ab] = \text{sum}(A, a, b) - \text{sum}(B, a, b)$ ， $val[cd] = \text{sum}(B, c, d) - \text{sum}(A, c, d)$
- 求 abcd，使 $val[ab] + val[cd]$ 最大（这题比较难，先写了个暴力解法，然后和面试官逐步讨论优化，没有给出最优解法）
- 汉字数字转数字，如 一百二十转化成 120

4.2.2 智力题

- 丢两个骰子，最可能出现的点数和是多少？3 个骰子呢，不能枚举，面试官让快速估计。
- 给定二维平面一些点，问用一个半径为 1 的圆最多可框住多少个点（说思路：每两个点确定两个对称的圆，当然两点距离不能超过直径 2）
- 给一个 $N \times M$ 的棋盘，从(1, 1)移动到(N, M)，只能向右或向下，计算方案数，如果 N 和 M 很大怎么办？

4.3 其他方面

4.3.1 概率分析

- 给一个随机函数 fun，30%的概率产生 1，70%产生 0，如何用 fun 产生等概率的 0 和 1？

4.3.2 矩阵运算

- 顺时针旋转正方形矩阵 90 度；在时间复杂度 $O(\log(m+n))$ 的要求下寻找两个有序数组合并后的中位数？

4.3.3 其他

- 给定一堆序列标号（标号用整数），形式为 $[a,b]$ ，表示标号为 a ， b 的两个物体的体积关系满足： $a > b$ ，用合适的数据结构存储数据，并判断这堆序列是否有效？
- KMP 算法有什么缺点？除了 KMP，还有什么算法可以快速做字符串查找？

4.4 Leetcode&剑指 offer 原题

- Leetcode 72
- Leetcode 137：写通用解法，要求时间复杂度 $O(n)$ ，空间复杂度 $O(1)$ ？
- LeetCode 283：原地移动数组，使得元素对应顺序不变，0 值移动到末尾
- LeetCode 543：二叉树直径
- LeetCode 1497：问复杂度
- Leetcode 原题：股票只能买入卖出一次，和买入卖出多次？
- LeetCode 原题：最大子数组和
- LeetCode 原题：实现 `atoi` 考虑所有情况，考虑所有异常情况，包括溢出
- 剑指 offer 原题：一个矩阵，每一行从左到右递增，每一列从上到下递增，给一个值，判断这个值是否在矩阵中？面试官后面要求用 `binary search` 做，问了下时间复杂度。

5 编程高频问题：Python&C/C++方面

第五节
微软面经
编程高频问题
(整理: 江大白)
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

无

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
微软面经
操作系统高频问题
(整理: 江大白)
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

- 数据库的索引有了解过吗，有哪些优缺点？

6.2 操作系统方面

- 什么是死锁，造成死锁的原因有哪些？

7 技术&产品&开放性问题

7.1 技术方面

- 场景题：如何给问答系统中的新问题推荐答案？
- 场景题：单词纠错怎么做？
- 场景题：如何让对话机器人产生的回答更具情感性，面试官简化了问题：机器人产生回答后，我们给回答加前缀，比如问“今天吃饭了吗？”，回答“【嗯呀】，我吃了”，如何从大规模 QA 数据中统计出要加哪些前缀（如上面的“嗯呀”），然后判断是否需要加前缀，需要加什么前缀？
- 场景题：对话机器人说了一句话后，如何判断该话是否含有***、暴力元素，有标注数据怎么做，无标注数据怎么做？

- n 个准确率为 50% 的分类器，可以通过什么方式提升准确吗？60% 呢？如果可以，提升到 96% 需要多少个？

7.2 产品方面

- 一些推荐算法相关，比如搜搞笑，推荐的视频如何排序，怎么区分推荐的视频是不是重复的等等？

24|oppo 算法岗武功秘籍

1 oppo 面经汇总资料

第一节
oppo面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 oppo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: oppo 面经-35 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【计算机视觉实习生】

(2) 全职岗位类

【机器学习算法工程师】、【NLP 算法工程师】、【视觉算法开发工程师】

1.3 面试流程时间安排

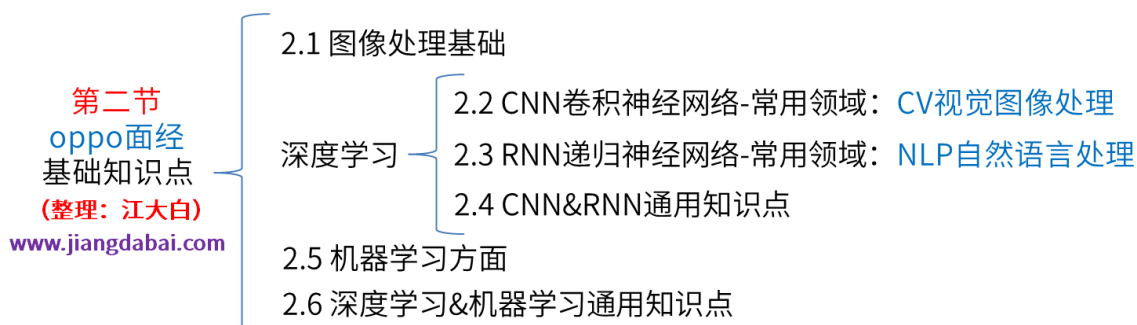
oppo面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	主要问一个个项目， 细节抠的很多
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	项目和基础知识
第三面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

1.4 oppo 面试心得汇总

- ★ 基础知识真的很重要。本人认为的计算机基础知识包括数据结构与算法，语言的掌握程度，操作系统原理，计算机网络，自己非科班这方面真的比较弱。这也是自己以后面试准备补充的知识。
- ★ 都是非常基础的问题，很考验基本功。项目和实习几乎没有问。
- ★ 会问非常基础的深度学习网络知识。 机器学习类的很多、推荐系统、广告系统之类 计算机视觉，比如目标检测之类的比较少。
- ★ Oppo 面试总结起来：面试确实有随机性，公司的风格和面试官的风格都有很大影响。 此外，综合面试和 HR 面的常见问题还是要准备一下的。

2 oppo 面经涉及基础知识点



2.1 图像处理基础

2.1.1 讲解相关原理

- 讲一下 LBP 算子？
- 了解 raw 格式图像不？
- 传统方法怎么进行图像模糊检测？

2.1.2 手写算法代码

- 手写中值滤波？
- 手写均值滤波？
- 手写直方图相似度计算？

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 卷积方面

- 卷积操作是线性的吗？CNN 是线性的吗？
- dropout 的原理，你为什么觉得 dropout 有效？

2.2.1.2 网络结构方面

- VGG, Googlenet, resnet 如何做网络轻量化的?
- Resnet 的原理讲一下?

2.2.1.3 其他方面

- 梯度消失和梯度爆炸的产生原因和解决办法?
- Batchnorm 如何解决梯度消失问题?

2.2.2 激活函数类

- Softmax 为什么 soft?

2.3 深度学习: RNN 递归神经网络方面

无

2.4 深度学习: CNN&RNN 通用的问题

2.4.1 基础知识点

- 深度学习过程中学习曲线产生震荡, 分析原因?

2.4.2 模型评价

- F1-score, AUC, 交叉熵; AUC 和交叉熵在分类上的区别与联系?
- AUC 及其几种用途?

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

- PCA 推导的原理？
- 稀疏矩阵如何处理，稀疏矩阵和低秩矩阵区别？

② 特征选择

无

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

- 介绍一下随机森林
- 说一下随机森林为什么比决策树好？
- 简单写一下随机森林的伪代码？

B.基于 boosting：Adaboost、GDBT、XGBoost

- GBDT 原理，XGB 原理，二阶泰勒展开怎么写，XGB 相比 GBDT 有哪些区别？
- XGB 怎么解决过拟合？怎么剪枝？如何选择特征？怎么处理缺失值？
- Adaboost, GBDT, XGBoost?
- xgb 讲一下、xgb 行列抽样、xgb 特征重要程度，你觉得他这样做合理吗？xgb 连续值你为什么要分箱？
- gbdt+lr 原理？
- gbdt 和 rf 区别，如何理解 gbdt 中的 gb，从偏差方差角度讲一下？

② 逻辑回归 LR

- LR 原理？推导 LR？
- LR 怎么加正则项？
- LR 和 SVM 的区别？（写出各自目标函数）

- 介绍 LR, gdbt 的数学原理

③ SVM (支持向量机)

- SVM 分类提的什么特征?
- SVM 介绍, 以及松弛变量的理解?

④ 决策树 (DT)

- 决策树, 信息增益公式, 基尼指数公式, 分类和回归用什么损失函数?
- 讲讲决策树? (讲了三种决策树及分支点选择策略: 手写信息增益和基尼系数等)
- 决策树怎么选择特征? 信息增益和信息增益比的区别?

2.5.1.4 无监督学习-聚类方面

- 列举已知的聚类算法

2.5.2 手推算法及代码

- LR 手推一下?

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

- 三元组损失函数?
- 交叉熵损失的优缺点, 为啥不用别的损失?

2.6.2 激活函数方面

- 常用的激活函数?

2.6.3 网络优化梯度下降方面

- 比较一下牛顿方法、SGD、GD、mini-batch GD?
- 说一下优化器, 你一般都用什么优化器。从动量法, adagrad, rmsprop, adam 说了一下。具体问了一下 adam 的公式。具体问了一下 adagrad, rmsprop 是如何对学习率做约束的?

2.6.4 正则化方面

- L1 正则和 L2 正则的形式？L1 在 $x=0$ 出不可导，怎么办？
- L1, L2 正则化的区别？具体使用场景？

2.6.5 过拟合&欠拟合方面

- 欠拟合，过拟合怎么回事，又怎么去解决？

2.6.6 其他方面

- 哪些指标不能用于正负样本不均衡，哪些可以，为什么？
- 特征向量和特征值分别有什么含义？

3 oppo 面经涉及项目知识

第三节
oppo面经
项目知识点
(整理: 江大白)
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

- Focal loss 介绍（项目用到）

3.1.2 图像分割

- Deeplab, psp, FCN 了解多少？讲一下 FCN 的具体结构。
- FCN 最后进行几层的拼合，为什么这么做？（卷积太深了，像素定位不准确。叠加拼合能够提高准确度。）
- unet 介绍（项目用到）

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

- Word2Vec 原理?
- Word2Vec 滑动窗口和维度设置有什么讲究么?
- fasttext 了解吗?
- tfidf 的劣势?

3.3 强化学习

无

3.4 机器学习方面

3.4.1 推荐系统

- 介绍 fm, ffm, deepfm, dcn, xdeepfm
- 问了 deepfm 是如何提取交叉特征的

4 数据结构与算法分析相关知识

第四节
oppo面经
数据结构与算法分析
(整理: 江大白)
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

- 连续子数组的最大和？

4.1.1.2 字符串

- 判断一个字符串是否是另一个字符串的子字符串(KMP 算法)？
- 求字符串中最多包含 K 个不同字符的最长字符串？
- 两个字符串，找到第二个字符串在第一个字符串中的下标，没有返回-1。直接让我用暴力解法做。
- 不使用字符串判断一个整数是不是符合回文，比如 121 就是满足要求。

4.1.2 树

- 二叉树了解吗？后序遍历是什么顺序？写一下代码。

4.1.3 排序

- 写一下冒泡排序？
- 求 TOPK 问题
- 写一个快速排序？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 海量数据查重选最高频率的 10 个？
- 问了一个列表里面有很多英文单词，怎样选出频率出现最高的前几个词？
- 打家劫舍问题

4.2.2 智力题

- 斗地主大小王分到同一个人手里的概率

4.3 其他方面

- 问项目，项目里用到了泊松过程，让写了一下泊松分布？

4.4 Leetcode&剑指 offer 原题

- Leetcode 221：最大正方形

5 编程高频问题：Python&C/C++方面

第五节
oppo面经
编程高频问题
(整理: 江大白)
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

- 堆和栈在内存分配上有什么区别？

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
oppo面经
操作系统高频问题
(整理: 江大白)
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

6.2.1 TCP 协议相关

● 介绍一下 tcp 解包的过程（这个不会，说了应用程序通过套接字吧数据下发给传输层，然后传输层进行传输，再通过另一个应用程序的套接字接收数据包）

6.2.2 其他问题

● 说一下用户态，内核态

7 技术&产品&开放性问题

7.1 技术方面

● 当遇到一个实际情况的时候，是一上来就用深度学习 cv 去解决还是先用传统图像处理方法，怎么去评估二者区别。做的项目中遇到了什么难点，怎么去解决的？

● 业务题：如何判断或者预测，食堂的就餐人数？

● 如果现在快递公司要在全中国建十个仓库，应该怎么优化？（k-mean）

● 深度学习了解吗？传统统计和深度学习你怎么看？

7.2 开放性问题

● 对人工智能应用的展望：这个我确实没准备，立马现想。首先讲了工业界和学术界的发展方向，然后讲了未来的应用，比如闲聊机器人（带情绪，和模仿某个偶像的）以及更加拟人化的智能助手（带情绪的，多轮对话的）。

25|Bigo 算法岗武功秘籍

1 Bigo 面经汇总资料

第一节
Bigo面经
汇总资料
(整理: 江大白)
www.jiangdabai.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 Bigo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: Bigo 面经-26 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 全职岗位类

【计算机视觉算法工程师】、【Bigo 推荐算法工程师】、【图像增强算法工程师】、【NLP 算法工程师】、【强化学习算法工程师】、【图形图像算法工程师】

1.3 面试流程时间安排

Bigo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	考察基础知识的掌握
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	项目+基础知识
第三面	总监面	自我介绍+项目/实习经验 +技术问答+算法编程	/
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：

- 有的是外地的，会用 zoom 进行远程面试，有的也可能三面都是电话面

1.4 Bigo 面试心得汇总

- ★ 感觉 bigo 的编程题跟之前面试的不太一样，比较开放性
- ★ 感觉基础还是问得比较多的，不过主要是基础，延伸不会很多
- ★ 问得还是比较全面的，python、C++、算法、项目，都会涉及
- ★ 个人感觉 BIGO 的面试体验还挺独特的，全程没有怎么问 DL，XGBOOST，GBDT，LR 这样的问题，倒是统计与优化这些问的特别深，我感觉非常有意思。
- ★ BIGO 笔试简单，但面试难度的话，就我所知的算法岗，个人认为不亚于任一大厂，用一个面 C++的朋友的话说，BIGO 面试问的非常硬核，很学院派，跪得很服气。而且周围拿到 offer 的不乏清北大佬、手握 n 篇（ $n \geq 2$ ）顶会 paper。
- ★ 签约以后，HR 给每个人分配了一个导师，一般是以后的 leader。和我对接的是前辈是中科院的博士，在雅虎 AILAB 和凤巢工作过 n 年，对于导师的资历还是很满意的。

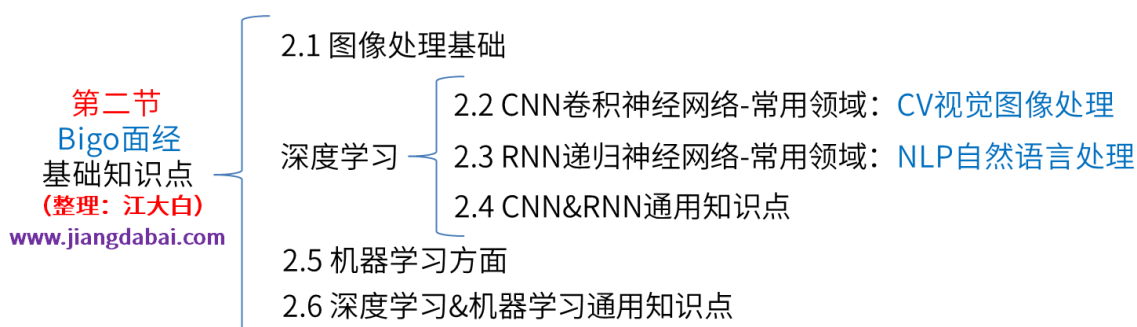
★ 有的人，在第一面之前，可能会有笔试，不过不太难。

(1) 编程题好像是 3-4 题，纯算法题均不超过剑指 offer 的难度，机器学习涉及到 AUC 的并行计算方式等等

(2) 概率题 1-2 题，例如算贝叶斯，红球白球等

(3) 机器学习基础 2，例如避免过拟合的方法等烂大街的

2 Bigo 面经涉及基础知识点



2.1 图像处理基础

- SIFT 的原理讲一下?
- 了解的图像纹理描述方法? (讲了 LBP 等)

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 卷积方面

- 卷积底层的实现方式，在 caffe 里这个方法叫什么名字? (im2col)
- 空洞卷积了解吗?
- dcn 可变形卷积讲一下?
- 亚像素卷积，一般上采样会有哪些问题?
- 上采样方法都有哪些?

- dropout 在训练和测试时不同，怎么保证测试结果稳定？
- dropout 的原理说一下？
- 感受野的计算/增大感受野

2.2.1.2 网络结构方面

- Vgg 和 Resnet 讲一下？ ResNet 的优缺点说一下
- 网络结构的演进，从 AlexNet 到 DenseNet？
- 各种分类网络简述，mobilenet 具体实现？
- 感觉他们用的还是手机端的多一点，问了 resnet 和 resnext 的发展历史和优缺点？(还得多准备一些)，我只是简单的谈了谈，并没有深入讲，还问了 mobilenet。

2.2.1.3 其他方面

- 批正则化 BN 的原理和作用，在训练和测试时有什么不同？
- BN 针对的是哪个维度？计算的维度？
- 梯度消失和梯度爆炸的原因？处理梯度消失的方法？
- 多任务学习中每部分损失函数的权重确定？
- Batch normalization 的原理？在训练和预测中的差异？

2.2.2 数学计算

- 卷积的计算量，depth-wise 的计算量？

2.2.3 公式推导

- 卷积后特征图大小的计算公式？
- 手写了三层神经网络的反向传播？
- 如果给定卷积核的尺寸 K，输入尺寸 WH，padding 为 P，步长为 S，空洞卷积参数 d=1，给出卷积之后的尺寸计算公式？（面试官将空洞卷积参数设为 1，其实就是常规卷积，降低了难度，重点是怎么计算空洞卷积实际覆盖的尺寸）

2.2.4 手写算法代码

- numpy 写卷积操作

2.3 深度学习：RNN 递归神经网络方面

无

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 数据不平衡怎么办？
- 训练的时候为什么要分成训练集、验证集？

2.4.2 模型评价

- 介绍 AUC，AUC 的范围，手写 AUC 代码？
- 评价标准，精确率，召回率讲一下？
- MAP 计算方式？AUC 怎么计算？
- ROC 曲线，他的横坐标和纵坐标？如果正负样本不均匀 ROC 曲线会变吗？
- 探讨了现在使用的图像评价指标的不合理性？

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

- PCA 了解吗？

- PCA 原理和执行步骤?
- PCA 作用, 奇异值分解?

② 特征选择

无

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)

- XGBoost 思想, 具体怎么做的, 怎么选取特征的?
- Xgboost 在多元分类与二分类过程中有什么不同? Xgboost 的原理?
- Boosting 是在对上一次训练结果的残差进行拟合, 为什么这么说呢?
- Adaboost/GBDT 分别拟合了什么残差?

② 逻辑回归 LR

- 逻辑回归怎么推导出来的?
- FM 模型的原理? 相较于 LR 的优势在哪里呢?

③ SVM (支持向量机)

- SVM 的具体实现机制, 整个流程?
- SVM 的目标, 用到的数学方法? (其中具体问了 KKT 条件是什么和为什么), 有哪些 trick? (其实是问的核函数), 还问了松弛因子和惩罚因子。
- 推导 SVM 对偶问题, 讲解 SMO 算法?
- SVM 拉格朗日乘子的原理? SVM 的软间隔?
- KKT 条件内容和原理?

④ 朴素贝叶斯 (Naive Bayes)

- 说一下贝叶斯优化? (各种抠细节, 从高斯过程回归, 说到贝叶斯优化, 到一些优缺点和改进形式, 每一步都很细)

2.5.1.4 无监督学习-聚类方面

- Kmeans 解释下?

2.5.2 手推算法及代码

- SVM 的数学公式以及推导过程?

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

- 写一下 RPN 的损失函数? (多任务损失:二分类损失+SmoothL1 损失)
- 推导交叉熵求导?
- 交叉熵和 KL 散度的联系?

2.6.2 激活函数方面

- Sigmoid 和 Relu 区别, 为什么现在都用 Relu 而不用 Sigmoid?
- Relu 和 Sigmoid 的优缺点说一下?

2.6.3 网络优化梯度下降方面

- 牛顿法的收敛速度, 是不是一定比梯度下降快? (举反例)
- 牛顿下降的几个 phase, KKT 的仔细描述?
- 梯度下降步长选择问题? (回溯法, 两阶段优化法, 梯度积累法)
- 除了梯度下降, 线性回归还能怎么求解?

2.6.4 正则化方面

- 常用的正则化方法有哪些?
- L1 和 L2 正则化的区别?
- L1 和 L2 正则化各有什么特点? 为什么会有这样的特点?

2.6.5 过拟合&欠拟合方面

- 缓解过拟合的方法?
- 树模型会不会过拟合，树模型防止过拟合的方法?

3 Bigo 面经涉及项目知识点

第三节
Bigo面经
项目知识点
(整理: 江大白)
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

- 介绍 Faster R-CNN 原理?
- RCNN->Fast RCNN->Faster RCNN 的内容?
- Faster RCNN, SSD, yolo 之间的区别?
- 写一下 RPN 的损失函数? (多任务损失:二分类损失+SmoothL1 损失)
- RPN 损失中的回归损失部分输入变量是怎么计算的?
- RoI Pooling 是怎么做的? 有什么缺陷?
- 目标检测有一个类别 AP 很低怎么办?

3.1.2 图像分割

- FCN 有了解过吗?
- 在人像分割中, 针对头发这种区域的分割, 往往处理不是很好, 怎么改进?
- 如何在分割中利用周围像素点进行建模, 提高分割结果?

3.1.3 超分辨

● 超分辨率问题设计深度网络上的注意事项和主要关注点，超分辨率问题和别的问题有什么区别？

● 从结构原理上对比超分最先进的网络特点以及优缺点？

● 为什么超分没有用 L2 loss？

● 超分辨率部分，自己做了哪些创新？

● SubPixel 的原理是什么？它的尺寸变换公式是什么？比如给定输入 shape: NCHW, subPixel 之后的 shape 是多少？（假设上采样因子为 4，即扩大 4 倍）

（这个地方跟面试官的观点有些冲突，我的答案是 $N(4C)(2H)(2W)$ ，面试官认为应该是 $N(C/4)(2H)(2W)$ 。最后他强调了一点，就是上采样不会增加参数计算量）

3.1.4 图形图像方面

● 渲染管线相关：

简单介绍渲染管线，从顶点着色器到片元着色器

● 顶点着色器和片元着色器分别用于做什么？在顶点着色器里主要进行什么操作？

● 光照的计算可以在顶点着色器进行吗？可以在片元着色器进行吗？两者有什么区别？

● 在顶点着色器中一般会对法线做什么修改？

● 顶点着色器主要进行哪些坐标转换？从世界坐标系中转换到模型坐标系中的变换矩阵是怎么得来的？

● 对于一个世界坐标系下有一个子空间（子空间三个坐标轴可能不会相互正交），子空间中有一个向量，如何将该向量变换到世界坐标系中？

● 片元着色器之后还会进行什么操作？

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

- Transformer 的原理？Transformer 在训练过程中有哪些可以调整的超参？Multi-head attention 的原理？

3.3 强化学习

- 针对我的强化学习推荐系统项目，问了强化学习如何做重排序问题？
- 场景题，用强化学习做推荐我会怎么做，从召回到精排整个流程设计，算法使用 DDPG？
- 强化学习推断的时候时间复杂度太大怎么解决？

3.4 机器学习方面

3.4.1 推荐系统

- 问了下 wide&deep 中有没有遇到预训练的 embedding，接到 deep 网络中由于前者参数的 scale 较大带来的问题？怎么解决？面试官说可以用两个网络，一个固定 deep 网络参数，一个固定 embedding

- 你对推荐系统的了解？

答曰：了解了推荐系统的主流框架，召回方法，排序模型，并简单介绍了基于 embedding 相似度的召回和 FM 模型。

4 数据结构与算法分析相关知识

第四节
Bigo面经
数据结构与算法分析
(整理：江大白)
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

- 从两个排序好的数组中找出中位数？（合并数组，分奇偶求中位数，也很简单）
- 数组中只有一个出现一次的数，其余的数出现两次，怎么找到这个数？衍生： 优化空间复杂度、时间复杂度，如果出现一次的数有两个怎么解决？
- 给一个数组，随机 shuffle 数组使得每个值在每个索引处出现的概率相等，也是给两个 API(`int random()`),返回 0-65535 的任意值或者 `int random(a,b)`,返回 $[a,b]$ 区间内的任意值)，最后还要证明自己的方法保证概率相等？
- 二维数组从左往右递增，从上往下递增 、写了 $O(M+N)$,面试官问有没有有 $O(\log(n))$ 的，二分的没答出来？

4.1.1.2 链表

- 链表循环右移 k 位？

4.1.2 树

- 二叉树输出最右节点？
- 每个二叉树节点都有一个高度，一个人站在右边向左看，输出这个人能看到的节点？
- 二叉树从根节点出发的最大路径
- 二叉树从任意节点出发最大路径

4.1.3 排序

- 千万级的数，选出最小的 100 个数！
- 寻找 N 个数中，第 K 大的数

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 螺旋矩阵的题目

4.2.2 智力题

- 一个硬币有一面比较厚，总是倾向于翻到这一面，利用这个硬币如何实现公平？

4.3 其他方面

4.3.1 数论

- Hessian 矩阵的求逆与计算量问题？（正则化，Hessian 矩阵的近似）
- 说一下运筹学和凸优化？（问的很细，说了梯度下降，牛顿法，KKT，单纯形，混合整数规划）

4.3.2 计算几何

- 给四个点的坐标，判断四个点是不是能够构成菱形？
- 一个单位球，均匀产生球表面的坐标？提示：用极坐标。

4.3.3 概率分析

- 算一个贝叶斯概率的问题，就是教科书上非常喜欢问的一个人诊断某种疾病阳性，问真实的得病概率？
- 一个圆，圆上有三个点，构成锐角三角形的概率？答案：1/4
- n 个红球， m 个白球，A，B 两个人轮流不放回取球，谁先取到红球谁获胜，求 A 获胜的概率？
- A，B 两人轮流和电脑下象棋，A 赢的概率 0.8，B 赢的概率 0.7，赢了的人继续下，输了换人，求最终 A 和 B 共同对电脑获胜的概率？
- 丢十个硬币，七个正面，用贝叶斯概率求正面朝上的概率？

4.3.4 矩阵运算

- 给一个矩阵 ($n*m$)，求出矩阵中每两行的 cos 相似度，放在一个 $n*n$ 的矩阵里，可以用两个 API (`matmul(a,b)`和 `reduce_sum(a,1)`) ?
- 矩阵顺时针打印?

4.3.5 其他

- $A-B+C$ ……一个表达式，假设所有变量都是 `int32`，求问交换运算符号会不会影响求值结果? `int32` 换成 `float`，会不会影响求值结果? 取具体的例子说明。
- 最大连续字序列的和与积，求和的说了一下思路，求积是手撕的?
- 从一个低字节的 `int` 型变量中分别提取每个字节的数据?
- 求属于最大子段积，是积不是和?
- 最大连续子列，要求能够保存多个相同的最优情况的起始点?

4.4 Leetcode&剑指 offer 原题

- Leetcode 53: 连续子序列的最大和
- Leetcode 358: 重排一个数组，让相同数字的间隔 $\geq d$ ，返回重新排列后的结果?

5 编程高频问题：Python&C/C++方面

第五节

Bigo面经

编程高频问题

(整理: 江大白)

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 网络框架方面

- pytorch 中 `train` 和 `eval` 下有什么不同?

- 熟悉和常用哪些框架？

5.1.2 基础知识

5.1.2.1 内存相关

- python 实现深拷贝？
- Python 的浅拷贝和深拷贝（字符串的拷贝）

5.1.2.2 区别比较

- python2 和 python3 的区别？(print, raw_input, xrange, 整除除法)
- mutable 和 immutable(后来有详细地问 `a="abc", a+="d" / a=[1], a.append(2)` 时的内存变化引导我)

5.1.2.3 讲解原理

- python 继承封装多态,线程和进程，多线程？
- 类的继承和基类
- 装饰器，@property
- python3 中基类怎么运作的？
- “==” 的具体实现机制
- python 中的数据类型？
- python 什么数据是可比较的，什么是不可比较的？

5.2 C/C++方面

5.2.1 基础知识

5.2.1.1 线程相关

- 线程间通信
- 进程和线程讲一下？

5.2.1.2 区别比较

- 左值引用和右值引用
- 指针传递和引用传递

5.2.1.3 讲解原理

- C++中虚析构函数的作用、static 的作用和特点？
- C++中什么是左值什么是右值？
- 有哪些线程锁？（生产者消费者模型）

5.2.1.4 讲解应用

- 一个类 A 派生出类 B，其中类 A 的析构函数不为虚，然后 new 一个子类对象并强制转换成父类，再 delete 这个对象会产生什么问题？
- 一个类 A 派生出类 B，A 中有一个虚方法，在 A 的构造函数中有调用该方法，此时 new 一个子类对象并强制转换成父类，会分别调用哪些函数？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
Bigo面经
操作系统高频问题
(整理: 江大白)
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

6.2.1 线程和进程相关

- 浅、深拷贝区别？

7 技术&产品&开放性问题

7.1 技术方面

- 写代码的时候，为什么要声明训练语句和测试语句？
- 结合 precision，求解自动驾驶中图像推导的多少秒发生一次错误？

7.2 开放性问题

- 你对自动驾驶的看法，对 REID 的看法等一系列开放性问题？

26|猿辅导算法岗武功秘籍

1 猿辅导面经汇总资料

第一节 猿辅导面经 汇总资料 (整理: 江大白) www.jiangdabai.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 猿辅导面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 猿辅导面经-32 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 面经框架&答案&目录&心得:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【猿辅导语音合成算法实习】

(2) 全职岗位类

【OCR 算法工程师】、【视频编码工程师】、【深度学习算法工程师】、【计算机视觉算法工程师】

1.3 面试流程时间安排

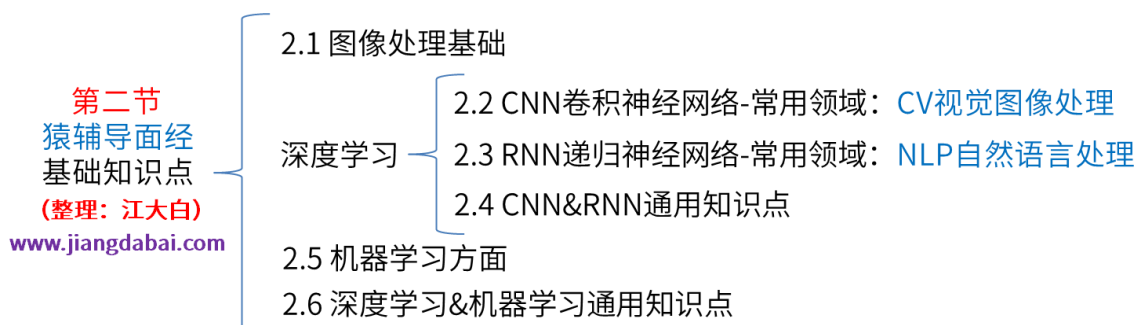
猿辅导面试流程 整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第三面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

1.4 猿辅导面试心得汇总

- ★ 一面手撕代码难度适中，面试官会不停引导；二面偏理论基础，考察了深度学习框架的使用、矩阵、概率；整个面试没有问太多项目细节。
- ★ python 编程的题虽然很简单，但是如果平时不多练几遍的话，真的会有很多细节上处理不好！
- ★ 很多都是 nlp、机器学习为主，计算机视觉很少
- ★ 项目要好好准备，和校招很大的区别是，面试官会问为什么做这个项目，前期的调研和数据支撑非常重要，这个问题回答不好的话，整个项目是没法让面试官信服的。
- ★ 还有项目中一些工业界常见的问题，前面提到的训练数据量不足的问题，还有模型训练时间，迭代周期的问题，如果迭代速度慢，怎么解决？
- ★ 关于方向的问题，工作一年还没有定型，所以不要担心换方向的问题，nlp 面 cv，推荐完全没问题。面试官更看重的是：基础扎实，工程实现能力强。
- ★ 猿辅导的福利薪资什么的不用说了，就很香。想加入猿辅导是因为视频领域在猿辅导这样的在线教育平台的应用真的是广泛而且多样，此外面试官提到的音视频实验室正在做的东西，也和专业很符合。

2 猿辅导面经涉及基础知识点



2.1 图像处理基础

- 搜寻矩阵中的连通域个数?

2.2 深度学习: CNN 卷积神经网络方面

2.2.1 讲解相关原理

- 介绍深度学习项目:

面试官主要的关注点在训练数据, 毕竟标注数据是稀缺资源。这类问题要着重准备, 即使面试官不问, 也可以主动提, 难点是标注数据太少了, 然后是怎么去做数据增广的

2.2.2 手写算法代码

- 实现一个 $m \times n$ 矩阵中 $k \times k$ 的 average pooling 算法, 不 padding, strides=1?

2.3 深度学习: RNN 递归神经网络方面

- LSTM 的原理讲一下?

2.4 深度学习: CNN&RNN 通用的问题

2.4.1 模型评价

- 模型效果怎么样; 用什么评价指标; 有什么可提升的方向?

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

- 数据集清洗做了什么？

2.5.1.2 特征工程

① 特征降维

- 矩阵分析中的奇异值与特征值的区别？

② 特征选择

无

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

- 随机森林有放回无放回，为什么？

B.基于 boosting：Adaboost、GDBT、XGBoost

- 有哪些常用 boosting 方法？
- adaboost 弱分类器要加权重吗，怎么加，gbdt 呢？

② 线性回归

- 线性回归的原理？

③ 逻辑回归 LR

- LR 的原理写一下？

④ SVM（支持向量机）

- SVM 的原理写一下？

⑤ 朴素贝叶斯（Naive Bayes）

- 朴素贝叶斯的原理讲一下?

2.5.2 手推算法及代码

无

2.6 深度学习&机器学习面经通用知识点

无

3 猿辅导面经涉及项目知识点

第三节
猿辅导面经
项目知识点
(整理: 江大白)
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

无

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

- HMM 维特比算法 (NLP 相关必考了, 这里的应用点是解决识别准确率)
- Beam search 的毛病, 返回很泛化的句子, 怎么解决?

3.3 强化学习

无

3.4 机器学习方面

无

3.5 视频编码

- 对 video codec 的理解
- 为什么想做视频编码？
- 编码框架（熟悉程度和运作流程）
- 编解码器技术细节
- 图像增强的应用

4 数据结构与算法分析相关知识点

第四节
猿辅导面经
数据结构与算法分析
(整理: 江大白)
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

- 找递增数组中第一个 $\geq n$ 的数字？
- 有序数组查找重复元素个数
- 找出来数组中每个元素后边第一个比它大的值？
- 二维数组回行打印？
- 一个无序有正有负数组，求乘积最大的三个数的乘积？
- 两个有序数组交集、并集？
- 数组插入排序
- 已排序的整数数组去重？

- 一个数组实现两个栈？
- 定长数组实现队列
- 无序数组构建一棵二叉排序树？
- 非降序数组，找与 target 最相近的数的下标？
- 任意一个整型数组，判断是否可以将数组分为三个区间，每个区间中数值的和相同，区间不能为空，时间复杂度要求为 $O(N)$ ？
- 单调不递减数组，给一个 target，找出大于等于 target 的下标 index？
- 不严格递增数组，要求删除出现次数大于 k 的数字，要求不要新建存储空间？
- 给定一个整数数组，每个数的范围为[0,100]，把这个数组分成两份，使得分割后的两个数组平均值相差最大

4.1.1.2 链表

- 链表 m 到 n 反转？
- 链表反转，分别用遍历与递归实现？
- 链表每隔 k 个反转？
- 双链表按照奇偶顺序分成两个链表，要求不要复制链表？
- 实现链表，无序链表，对链表值奇偶分离并排序，空间复杂度 $O(1)$ ？
- k 路链表归并？
- 链表实现队列
- 给定一个有序存在重复的值链表，使得每个元素只出现一次？
- 单调不递减链表，删除掉重复值？
- 链表相邻元素交换？
- 两个数字被逆序按位保存成链表，实现两个链表相加保存到第三个链表，例：
3->9->9->1+4->9->1？
- 输入链表 453612, target 3。输出 451236，就是把 target 后面的小于 target 的数移到 target

前，其余都保持相对关系，返回链表头节点

- 链表删除奇数位置的元素

4.1.1.3 栈

- 写一下栈排序
- 两个栈实现队列？
- 用一个队列实现栈

4.1.1.4 队列

- 队列实现栈（pop、top、push、is_Empty）

4.1.1.5 字符串

- 字符串形式自定义进制大数相加？
- 对字符串"ABCDE"按字母顺序的全排列，求 BDACE 排在全排列的第几个？并问该题的时间复杂度？
- 字符串横向改纵向？
- 字符串全排列，可能有重复的，要去重？
- 一些数，任意排列求可形成最小的值？
- 字符串中找出一个最长的最多包含 K 种字符的子串，输出长度，例如 s=“ebecda”，k=2。最长“ebe”，输出 3。
- 给定一个字符串如“aaabbcdade”，输出“a,3|b,2|c,1|d,1|a,1|e,1”

4.1.2 树

- 二叉树是否对称？
- 之字形打印二叉树，输出节点的值？
- 一个图，给定几条边，每条边有两个节点。已知它并不是一个二叉树，请判断去掉哪条边之后该图是二叉树？

- 无序数组构建一棵二叉排序树？
- 完全二叉树的最大深度与节点个数？
- 二叉树的最小公共祖先？
- 二叉树逆时针打印最外层节点？
- 输入一个数字 n ，构建一个完全二叉树并输出？
- 数字 n ，生成完全二叉树 1, 2, 3, 4,..., n ？
- 给定一棵二叉树，请找出其最长路径，并输出路径上的边的数量。如下所示，第一棵树的最长路径为 4-2-1-3，长度为 3。第二棵树的最长路径为 8-6-4-2-5-7-9，长度为 6？
- 二叉树宽度？
- 求二叉树的深度，不使用递归？
- 根据前序和中序遍历序列恢复二叉树
- 输出根节点到叶子节点路径之和为 target 的路径列表？
- 打印出根节点到叶子节点的最长路径？
- 第 K 层叶子节点个数？
- 判断一颗树是不是二叉搜索树？
- 从二叉搜索树到更大和树？
- 二叉搜索树转有序双向链表？
- 二叉搜索树第 k 个节点，不用中序遍历？
- 输入二叉树，target，输出 所有从根节点到叶子结点的和为 target 的 path

4.1.3 排序

- 手撕快排？
- 实现一个不考虑转移概率的维特比，要求给出 topn 的路径（现场写代码，用了堆实现的，没太准备好，写的比较乱）

4.2 算法思想实战及智力题

4.2.1 算法思想实战

● 给定一个 int 类型的列表（如[1, 9, 8, 8]）和一个加和元素（如 34）,计算列表元素所组成的数字（1988）与加和元素相加后的结果（1988+34=2022，此处仅作为讲解示例，实际操作中不允许这样执行），并以列表的形式输出（[2, 0, 2, 2]）？

● 对于普通的孤岛问题，求矩阵中可以经过岛屿到达边界的 1 的个数？

● LeetCode 773: 和 773 类似，滑动谜题

给定 2*3 棋盘的初始状态如

[[0, 1, 2],

[3, 4, 5]]

结束状态如

[[2, 1, 0],

[3, 4, 5]]

每一步可以使 0 与相邻数字交换，问从初始状态到结束状态最少交换多少步，若不能到结束状态返回-1

4.2.2 智力题

● 判断五子棋输赢，不用考虑复杂度？

4.3 其他方面

4.3.1 数论

● 用二分法对一个数字开根号？

4.3.2 矩阵运算

● 行和列都是有序的二维矩阵找一个 target 值？

4.4 Leetcode&剑指 offer 原题

- Leetcode 33: 搜索旋转排序数组
- Leetcode 200
- LeetCode 340: 至多包含 K 个不同字符的最长子串
- LeetCode 582: 杀死进程, 用了 DFS 解决
- LeetCode 688: “马” 在棋盘上的概率
- LeetCode 827: 最大人工岛

5 编程高频问题: Python&C/C++方面

第五节
猿辅导面经
编程高频问题
(整理: 江大白)
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

- 理解 Tensorflow 内部求导的原理吗?

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
猿辅导面经
操作系统高频问题
(整理: 江大白)
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

无

7 技术&产品&开放性问题

7.1 技术方面

- 数据流新加入一个数据之后如何更新均值、方差，推公式？