

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-59 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 面经框架及参考答案:

- (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

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

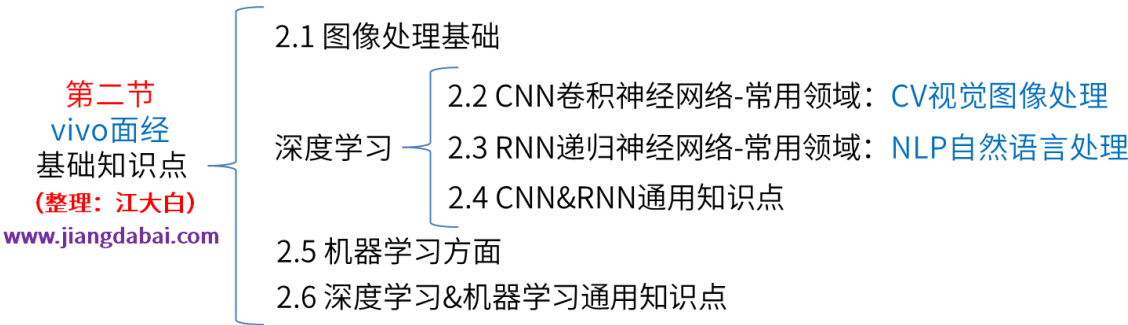
Vivo面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答	针对项目， 问改进点，创新方式
第二面	HR面	基础人力问题	/

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，用了的）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、GDBT、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？

4 数据结构与算法分析相关知识点

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？