

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-51 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

【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：

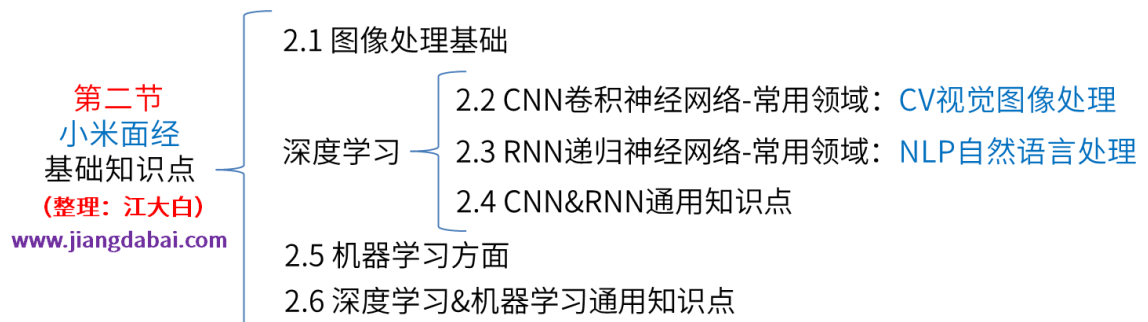
- 博士只有两轮，硕士可能会有三轮技术面

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 的参数数量和 FLOPs 计算公式？
- MobileNetV2 module 的参数数量和 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 gbdm 区别？ rf 和 gbdm 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

- gbdm, xgboost 模型的比较？ gbdm 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？
- gbd 的梯度为什么能代替残差？
- gbd 的损失函数是什么？
- gbd 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) gbd 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点

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, ..., 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= $a*b$ /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？