

34|蘑菇街算法岗武功秘籍

1 蘑菇街面经汇总资料

第一节
蘑菇街面经
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 蘑菇街面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 蘑菇街面经-13 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

【图像算法实习生】

- (2) 全职岗位类

【应用算法工程师】、【图像算法工程师】、【推荐算法工程师】、

1.3 面试流程时间安排

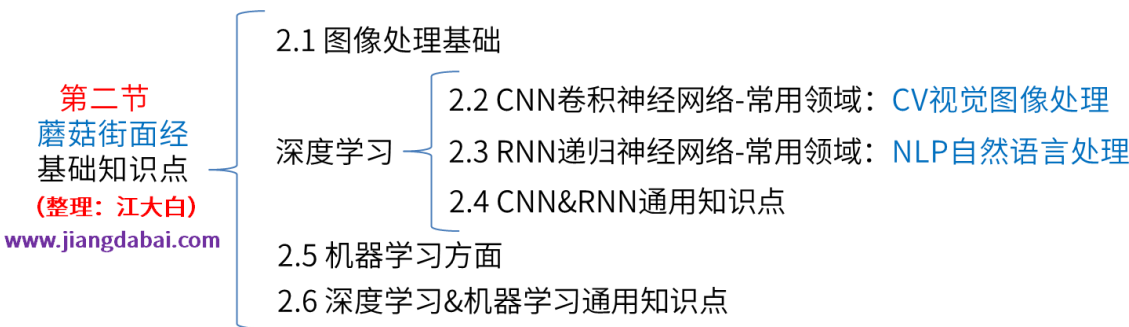
蘑菇街面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第三面	技术Leader面	自我介绍+项目经验+公司发展	/
第四面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

1.4 蘑菇街面试心得汇总

- ★ 面试官人很好，说什么都会给你回应。对项目会问很多细节。
- ★ 聊项目，问了类别不均衡问题，也问了模型可解释问题，好像蘑菇街喜欢可解释强的模型？然后当时引申的一个问题是，如果 nn 可以解决高维度离散问题，为啥还需要 gcn 来解决关联性的问题。其它不记得了。

2 蘑菇街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 网络结构方面

- 图像分类基础网络有何不同改进？

2.2.1.2 其他方面

- BN 层的均值如何训练的？

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 有什么不同？
- LSTM 如何解决 RNN 中梯度消失的问题？
- 项目中分别使用了 LSTM 和 CNN 进行文本分类，效果比较，为什么项目中 LSTM 的分类效果好？问了 RNN、LSTM 和 CNN 使用场景？
- 什么是梯度消失？LSTM 为什么能解决 RNN 梯度消失问题？

2.3.2 手绘网络原理

- 手写 LSTM 门控公式？

2.4 深度学习：CNN&RNN 通用的问题

2.4.1 基础知识点

- 方差和偏差是怎么回事？

2.4.2 模型评价

- 知道 AUC 吗，不均衡会对 AUC 产生影响吗，为什么？

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

- 做过哪些特征工程，产生哪些类型特征(统计特征，交叉特征，转化特征都有什么)，异常值怎么处理，数据平滑问题？

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 调过哪些参数？
- 正负样本比对 lr, gbdn 影响，为什么，用 lr 比例 N/P 比为 1 会怎么样？

② 逻辑回归 LR

- LR 和 GBDT 有什么区别，为什么 GBDT 结果比 LR 好？
- 正负样本比对 LR，gbdt 影响，为什么，用 LR 比例 N/P 比为 1 会怎么样？
- LR 原理、损失函数？
- LR 的详细过程怎么处理？
- 你说 LR 模型可解释性高，它的可解释性体现在哪里？是权重越大，可解释性就越强吗？

③ SVM（支持向量机）

- SVM 原理、损失函数？

④ 朴素贝叶斯（Naive Bayes）

- 贝叶斯公式的原理？

⑤ 决策树（DT）

- 你说一下信息增益的原理？

2.5.1.4 无监督学习-聚类方面

- 详细说一下 K-means 聚类，你每个簇怎么选样本的，距离用什么算的？

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

2.6.1 损失函数方面

- 知道哪些损失函数？

2.6.2 网络优化梯度下降方面

- 知道熟悉什么优化算法，SGD 和 Adam 的公式是什么样的，怎么迭代计算的？

- Adam 的思想?Adam 能否适用于稀疏数据?
- SGD, Adam 等, 以及他们的参数更新公式?
- 梯度下降如何跳出局部最优值?
- 如何解决陷入局部最小, 自适应学习率方法?

2.6.3 压缩&剪枝&量化&加速

- 模型简化压缩怎么处理的?

2.6.4 过拟合&欠拟合方面

- 过拟合的原因? 如何解决过拟合?

2.6.5 其他方面

- 机器学习和神经网络的选择问题? (什么情况下使用神经网络会取得较好的效果)

3 蘑菇街面经涉及项目知识点

第三节
蘑菇街面经
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

- Soft-NMS 和 NMS 的不同点?
- FPN 的原理?

- 项目细节，检测算法目前是怎么做的，跟踪问题经典的一些方法以及目前比较常用的方法，以及你的创新改进？

3.1.1.2 损失函数

- RetinaNet 的 loss function 是什么？
- 用到的 Focal loss 的公式，怎么用的，原理参数含义？
- Focal loss 是怎么缓解样本不均衡问题的，还有其他的方法吗？

3.1.1.3 手写代码

- 手写 NMS？

3.1.2 目标重识别

- ReID 用的网络是什么样的，损失函数是怎么构建的，公式是什么样的？

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 HMM 隐马尔科夫模型

- HMM 的维特比算法

② Word2vec

- 介绍 Word2Vec，我就把把 CBOW, skip-gram, hierarchical softmax, negative sampling, 介绍了一遍？

③ CNN 方面

- 介绍 TextCNN？

3.3 强化学习

3.3.1 讲解原理

- 问 Q 方法，说一下强化学习中 DQN 算法？
- 强化学习的分类？
- 值函数方法有哪些？
- 随机策略相对非随机策略优缺点？
- 熟悉的策略梯度方法？
- 如何改进 ppo 算法？
- 对 V 函数和 Q 函数的看法，为什么要有 V 和 Q？

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
蘑菇街面经
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

- 最大连续子数组和？

- 两个排序数组合并成一个数组?
- 给一个数组和一个数, 求数组中有多少对和为该数的数?

4.1.1.2 链表

- 实现链表反转?
- 两个单链表并集?

4.1.1.3 字符串

- 大数相加, 转换成字符串, 对每个字符相加减, 考虑进位?
- 字符串排列

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 分糖果问题?
- 10 个大文件返回词频 top10 的词?
- 统计 top100 词频?

4.3 其他方面

4.3.1 矩阵运算

- 写一下矩阵相乘代码, 我有三列是数据, itemid category ctr, 我想取每一类的 item 里面 ctrctr 最高的 100 个, 写一下代码?

4.4 Leetcode&剑指 offer 原题

- Leetcode 135: 分糖果问题

5 编程高频问题：Python&C/C++方面

第五节
蘑菇街面经
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 的装饰器解释一下？

5.2 C/C++方面

- C++动态绑定，数据结构了解不？

6 操作系统高频问题：数据库&线程&常用命令等

第六节
蘑菇街面经
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 产品方面

● 场景题：比如用户点击了广告跳转进来，要展示一屏幕的物品给他，该怎么做？可以说一下整个流程吗？

● 一个商家。第一天看 A 商品和 B 商品的点击率。点击率公式是点击/曝光。然后，平均的是 $\frac{a \text{ 的点击} + b \text{ 的点击}}{a \text{ 的曝光} + b \text{ 的曝光}}$ 。问，第二天发现，a 的点击率和 b 的点击率上升了，但是平均的下降了。是什么原因，从哪些方面入手去解决这个问题？