

腾讯、百度、滴滴最新NLP算法面经

原创 ZJU跑男F3 NLP情报局 1月18日

收录于话题

#面试经验 1 #自然语言处理 5



文 | ZJU跑男F3

编 | NLP情报局

因多方面因素，2020年的秋招形式依然严峻，特别是AI相关的算法岗位，竞争十分火热，既内卷又倒挂。

为帮助读者在即将到来的2021春招中斩获心仪offer，我们将整理一系列求职文章（详情可关注公众号“NLP情报局——乘风破浪”专栏）。

本文介绍腾讯、百度、滴滴三家大厂NLP算法的最新面试经验，其中前两家公司给了ssp offer。

腾讯TEG实习

视频一面

1、自我介绍

自我介绍即开场白需要提前准备并多次练习，内容不用太多，言简意赅地讲清楚**学校、专业、研究方向，概括介绍实习/论文/比赛等亮点**即可，细节可以根据面试官的兴趣再深入。

2、项目/实习经历

这部分主要根据简历上的实习/项目进行，可能会问背景、怎么做、为什么这样做、最终结果、是否成功上线等，再深挖一些技术细节，或顺带考察ML、DL等知识点。



由于校招同学没有工作经验，所以**项目和实习**是展示自身能力的最好亮点（一堆CCF-A的大佬请忽略并接受我的膝盖）。面试官可以从介绍中了解你的工程能力、抗压能力、沟通能力、思维方式等等，如果吸引到面试官肯定能加分，要是面试官觉得索然无味估计面试也就到这儿了...

所以**对于简历上的内容要滚瓜烂熟**，能知其然知其所以然最佳。如果问到简历上的内容没答上或吞吞吐吐，将非常打脸，造成硬伤（简历上自己写的都不会，别的还用问吗、(¬_¬)）。

你奈我何？



切忌在简历上洋洋洒洒地写“熟悉Xgboost、SVM、Bayes、HMM、CRF、KNN、LR、CNN、Attention等算法”。是听过名字就算熟悉？还是啃过源码、理解背后原理？即便真的对这些算法理解透彻，最好也能结合例子来说明，而不是穷举。

而对于没有项目/实习/比赛经历的同学，这一部分时间只能由考察基础算法来弥补了，这样一来随机性、难度都提升了不少；所以平时尽可能多积累这方面的经验，让简历充实起来。

“项目 or 实习 or 比赛 or 论文”最好为真，这是底线。

3、介绍一个比赛

我从简历上3个比赛中，选择了一个最有亮点的比赛，从数据、模型、提分trick三方面做了介绍。

4、如何实现知识蒸馏，loss怎么选择

5、介绍TinyBert的工作原理

6-9考察机器学习基础

6、LR是线性模型吗，为什么？

7、梯度下降和随机梯度下降的区别

8、GBDT和XGBoost的区别（至少3方面）

9、偏差和方差的区别

10-11考察深度学习基础

10、怎么理解XLNet的，XLNet的输入是什么？

11、一个二分类任务，假设只有一个维度的特征，取值范围是0 ~ 正无穷，如何实现二分类？

12、手撕代码：两个数组求交集

只能使用 list，如果要用其他数据结构（比如字典）需要自己实现；时间复杂度尽可能低。

腾讯一面难度中规中矩，量比较多但基本都在考察基础。

关于手撕代码，今年应该是非常关键也是必做的环节，很多公司直接将编程题作为通过面试与否的硬性指标。刷题要趁早！

电话二面

1、询问两个实习项目，详细考察细节

2、介绍textcnn实现过程，池化时loss如何反向传播

3、假如不用bert，一大堆文本有几百万个词，直接训练词向量很困难，如何去做特征选择

4、LSTM门控机制是怎样的

5、LTSM主要解决了RNN什么问题，为什么能解决

6-9重点考察语言模型

6、介绍GPT2如何写诗、写对联

7、GPT2如何围绕一个主题/关键词写诗

8、ELMO、BERT、GPT模型彼此之间有什么区别

9、word2vec和ELMO主要有什么不同，为什么elmo效果更好

hr视频面

1、询问导师、实验室的研究方向

2、智能客服系统中，你主要负责哪些模块？

3、base在深圳能否接受，多久可以来实习？

面试结束后过了2天，成功收到TEG的offer call。

百度自然语言处理

视频一面

1、手撕代码：

给定一个字符串S和一个字符C。返回一个代表字符串S中每个字符到C的最短距离的数组。

```
1 # input
2 S = "paddlepaddle", C = 'd'
3 # output
4 [2, 1, 0, 0, 1, 2, 2, 1, 0, 0, 1, 2]
```

2、介绍简历上的2个项目和1个比赛

3、介绍bert和TinyBert

- 4、bert如何实现微调
- 5、fasttext的实现原理
- 6、文本匹配如何实现
- 7、手推信息增益、信息增益率、Gini指数公式
- 8、ngram如何过滤不合理的组合
- 9、还有什么想和我分享的？

面试官给了一个展现自我的机会，我介绍了一个最近在实验室新做的项目（简历上没有）。

视频二面

- 1、手撕代码：

word2vec负采样具体怎么做的（知识点：累积概率分布、二分区间查找）

- 2、文本匹配怎么做
- 3、文本匹配不当分类来做，loss怎么设计？
- 4、bert在中文和英文场景下使用，有什么区别
- 5、介绍一下在开源组织CLUE的工作
- 6、假如给offer，什么时候能来实习

hr视频面

聊个人特质、看重公司哪些方面、个人职业规划等

滴滴出行规划

视频一面

- 1、详细介绍1个项目和1个比赛

面试官让现场打开比赛链接查看排名和解决方案

2、手撕代码：

- 因式分解
- 判断年份是否为闰年

视频二面

1、介绍1个项目

2、介绍transformer

3、介绍常用的聚类算法（KMeans、DBSCAN、Mean Shift）

4、介绍word2vec2种实现模型、区别以及和fasttext的区别

5、手撕代码：并查集

视频三面

1、介绍1个项目

2、手撕代码：

两个有序数组，求第k大，时间复杂度尽可能低（ $\log(m+n)$ ）

总结

近两年AI算法岗非常内卷，加上疫情影响给找好工作都增加了难度。但身边的朋友基本都在头部大厂收获了1+offer。实习和项目经验仍然很重要，ACL、NIPS等论文近些年在贬值。扎实的理论基础和coding能力可以让你进大厂后继续保持足够的竞争力。

手撕代码是面试的一道坎，要早做准备。同时，尽可能把握住可以跳过笔试（大部分公司）的**提前批面试**。

祝大家在新的一年里offer多多，好运连连。

- END -