

别再搜集面经啦！小夕教你斩下NLP算法岗offer！

原创 夕小瑶 夕小瑶的卖萌屋 2018-10-03

来自专辑

卖萌屋@互联网算法岗求职

>

推完上一篇文章，订阅号和知乎后台有好多小伙伴跟小夕要面经（还有个要买简历的是什么鬼），然鹅小夕真的没有整理面经呀，真的木有时间(。 ㄋ ｡)。不过话说回来，面经有多大用呢？最起码对于NLP岗位的面试来说，小夕发现根本不是面经中说的样子。。。

其实今年参加NLP算法岗秋招的小伙伴可能有感慨，
“照着别人的面经去准备了辣么多，轮到自己面试时内容怎么这么不一样？！”
“说好的要做到熟练推导SVM呢？怎么从来没人让我推导SVM？”
“整场面试都在聊前沿论文啊什么鬼？从来没见过这样的面经呀！”
“为什么面试官只问他简历内容，轮到我时一点简历内容都不问？”

不止今年，将来的面试肯定更趋向于千人千面，毕竟AI行业也将越来越细化嘛。所以小夕精心准备了这篇“万能”文章给你们，希望大家将来都能收割到自己想要的offer～

这篇文章不是面经集合，也不是装X和贩卖焦虑的晒offer贴，也不是堆砌可能问到的知识点，而是希望给还在迷茫的小伙伴提供一些系统的指导和建议。当然，这些建议可能不适合每个人，因此希望大家从这篇文章里汲取到适合自己的养分，千万不要邯郸学步哦m(——)m

目标观众

首先，本文的讨论范围仅限自然语言处理（NLP）相关岗位！亲测NLP岗的套路跟数据挖掘、机器学习岗还是有明显差异的，所以千万不要按本文的套路去准备DM、ML的面试！

然后说一下本文的目标群体。想去拿Google AI中心、FAIR、MSRA offer的大佬请放过小夕(。 ㄋ ｡)真的木有经验可以分享。还有那些已经一大把顶会的小伙伴，求给小夕指导好么！

其实之前小夕也曾想着去外企的研究院（虽然菜的一无所），然而很扎心的一次次打听到外企的研究院基本只招PhD。虽然外企的业务部门也有NLP岗位，而且硕士也能进，但是进一步打听了一下，听说硕士进去很少有research相关的工作，最多做做模型优化，弄不好就是爬数据、清洗数据了。后来拿了一家的offer后发现确实如此，加上后来没什么面试的状态了，也就放弃这一块了。不过有条件的同学可以在外企研究院或业务部门做做intern，这方面对硕士友好的多，而且听说很多业务部门的research工作也会让intern来做。

言归正传，对于想拿国内一线大厂的核心研究部门和核心业务部门NLP算法岗/研究岗offer的童鞋，转发本文，你就收获了半个offer！打赏本文，你就收获了0.75个offer！（等等，文风好像越来越像传销了。。。）

国内的NLP方向的岗位一般叫做NLP研究员/算法工程师/工程师，如果你的目标是跟下面小夕的战绩一样，那么本文或许可以给你提供点启发。

按流程来说吧。首先简历投了一圈都很幸运的木有沉（算了算差不多十来家，很多没内推）；笔试参加了网易、网易游戏和今日头条，都很幸运的木有挂（虽然做完后都感觉要挂）；面试过程中跪掉了腾讯和头条（小夕辣么可爱，怎么能忍心拒绝(。 ㄋ ｡)），最后拿了下面几个offer：

百度和阿里的SP offer或以上（某厂offer我不确定是不是该叫SSP），还有网易游戏等几个第二三梯队、快手等startup公司的offer（出了等级的也都给了SP及以上）。

ps：其他公司都是各种原因没去笔试或面试（记错了时间或者时间冲突，HR又不给调时间m(_ _)m)

好了，如果确定你是本文的目标读者，欢迎往下看。

Research or Product ?

首先，要搞懂自己想要什么。粗略的看，如果想做research，比如发发paper，打打国际比赛，做做通用技术平台，为现在或将来的产品孵化算法接口，就去偏research的部门；如果想直接优化产品的算法指标，看到自己的成果快速反映在产品的用户体验和经济效益上，那就去偏业务的部门。这两种不同导向的部门在面试的时候的关注点会存在一定差异。

当然，再偏research的部门也会把成果直接或间接的输出到各个产品线上（否则养你干嘛），再偏业务的部门也会有做research的团队，毕竟NLP本身还没有成熟嘛。

不过总的来说，比如你想做research，有不错的论文或比赛经历，对某个方向研究的比较深入，但做系统的经历相对匮乏，则偏research的部门可能更赏识你；如果你想做产品，并且做系统、啪代码的能力不错，论文也刷了不少但是科研热情不高，则你可能跟业务部门更match。

说了这么多，到底该怎么判断一个部门偏research还是偏业务呢？最靠谱的还是跟HR或者该部门的师兄师姐聊，如果没有，可以根据名字猜测一下，比如下面给出几个case：

ps：以下case跟人家真正的定位没有半毛钱关系，信息多总结自小夕和小伙伴的面试主观感受，注意是主观感受

偏research/通用平台的：

- 百度：自然语言处理部
- 阿里：达摩院的数据科学研究院（IDST）
- 腾讯：AI lab
- 网易游戏：伏羲AI实验室
- 京东：AI lab
- 滴滴：AI lab
- 商汤：研究院（没错，商汤研究院也有NLP）

偏业务/产品的：

- 百度：度秘事业部、智能客服部
- 阿里：智能服务部（阿里小蜜）、达摩院的AI labs（天猫精灵等）
- 网易：有道事业部

好了，搞懂了自己想做什么，瞄准一两个目标部门立个flag，后续就要围绕这根主线来哦。有条件的童鞋可以找师兄师姐提前打听一下意向部门的研究方向或者业务方向，免得准备了一年后发现人家根本不做这个。。。如果实在打听不到，可以搜罗一下该部门近几年的顶会paper、专利、宣传文章之类的，结合官网上的宣传，基本可以摸个差不多。

啰嗦了好多，下面就分享经验啦。内容分私货篇和干货篇，私货篇主要讲前期的准备（笔试、简历、论文、比赛、实习等，没耐心的同学可以只看加粗的内容），干货篇看似是对小夕几十场面试的面经的知识点浓缩，实际上只是为了说明一个trick，即别人的面经不重要，去为自己的关键词去量身定做复习计划吧。文末给出一些超级有信息量的tips。

私货篇

1. 笔试与面试中的笔试的准备

这方面小夕走了不少弯路，其实提前批（7月下旬到8月底）基本不会有笔试（今日头条、网易、网易游戏这三家有），因此大可把主要精力放在简历及延伸内容里。但是！解基本编程题的能力还是要锻炼出来的，这是红线。

面试中的白板编程题一般都超级简单，基本都是leetcode简单级别的或者剑指offer上的原题，偶尔出个leetcode中等难度的题。小夕把剑指offer刷了不到一半，leetcode刷了30来道题（按类别和top review过滤后刷，先刷简单的），发现应付提

前批面试中的编程题基本都可以应付了。当然，精力允许的话还是多刷点，毕竟解题解得快，别留边界问题，会给面试官印象好一些。

哦对了，编程语言的话，放心的用python就好，小夕实在没时间复习C/C++/Java了，所以leetcode还有各家的笔试面试都是用python写的，亲测没毛病。

至于计算机网络、操作系统、计算机组成原理、数据库、编程语言、设计模式之类的基础课，闲暇时翻一翻吧，全都从头复习一遍太费精力了，这些一般只出现在笔试的选择题中。面试过程中除了今日头条，其他家从来没考过。

另外，矩阵论、概率统计这些数学课也记得顺手复习哦，听后来参加正式校招的小伙伴说，有的大厂的算法岗试题不怎么考计算机基础，本科的数学课倒是考了很多。不过小夕亲测硕士阶段的数学课像最优化之类的，考的非常非常浅，科普级别的深度就够应付了。（感觉最优化白学了(´Д`)

最后小夕不靠谱的猜测有的厂的笔试刷人可能不全看总分。因为小夕在网易的笔试的选择题部分瞎蒙了好多题，最后40分的问答题也完全空着没做，就中间30分的编程题磕磕绊绊的基本AC了（也挂了几个case），但是最后笔试竟然过了。

2. 简历的准备

翻了一下，发现自己的简历前前后后改了23版。。。回看7月份的简历，简直辣眼睛(╯▽╰)

毫无疑问简历是最最重要的东西，它不仅是面试的敲门砖，而且基本贯穿了整个面试过程！从小夕的经历看，不加面试中的白板编程时间，简历内容一般会占据每轮面试80%-100%的时间（今日头条除外），其中占到100%时间的面试能占到50%以上。

做简历的最基本也是最关键也是最容易被忽视的一条原则是，整体内容一定要跟岗位需求match！其实做简历跟写文章很像，要围绕一根主线展开，如果发现电路也做，视觉也做，推荐也做，NLP也做的话，很容易被打上“跟岗位不match”或者“这孩子做事浮躁”的tag。（身边一小伙伴就这样被某二梯队厂以“NLP经历不足”为理由挂掉）

对于已经一串paper的那种简历来说，就不做建议了。但是如果对于比赛、论文、项目、专利都只有几条的童鞋来说，完全可以把简历做的有的放矢，match的内容展开写，不match的一笔带过或者直接不写，这样面试官也省的一条条甄别信息量，面试过程也会变得更加主题明确。

小夕的一个小trick是，可以尝试把最有信心在面试中谈起的经历所在的板块写在最前面（仅次于教育经历），并用配色突出这条经历。小夕亲测这样的第一条经历可以聚焦非常多的面试火力，甚至有15%的面试是全部面试时间都在死磕这条经历！而且小夕神奇的发现，各轮面试很少有第一轮面试官问第一条，第二轮面试官问第二条这种，倒是经常有面试官觉得你第一条经历很亮，于是告诉下一轮面试官这条很亮，然后下一轮面试官出于好奇就深入问这条经历，如果他也觉得很亮，就会转述再下一轮面试官，于是再下一轮面试官又会主要问你这条经历(╯▽╰)，然后offer就到手了。当然，这样的副作用就是，如果对面试官来说这条经历不亮，那面试基本挂掉一半了。

另外，除非你硕士期间对一个方向钻的很深、已经可以驾驭这个方向的绝大部分问题了，否则小夕墙裂建议单独设置一个板块列一下自己研究过的算法问题。比如分两级，第一级讲方向（比如对话系统），第二级讲具体研究的子问题（比如聊天的一致性问题），这样可以避免面试官对你进行天马行空式的考察，毕竟硕士时间有限，哪怕只做NLP中的一个方向，也难以研究的面面俱到。该板块会贡献大量的关键词，这些关键词基本可以决定面试官对你的考察范围，详见后面的干货篇。

还有一个trick，简历内容太多，做到两页后，难免最后要塞一些无关痛痒的荣誉（比如奖学金啥的）还有一些主观内容（比如算法研究、IT技能、自我评价之类的），这样面试官看到后面发现没营养了反而容易忘掉前面的亮点，所以小夕是把一条亮点经历放在简历最后点个题（记得颜色突出一下，否则真被忽略了就哭了），这样会让面试官觉得整个简历“干货满满，这么亮的经历都只能排到最后面了”。

最后，对小夕简历好奇的小伙伴也不要好奇了，除了性别没什么亮点(╯▽╰)各种被HR指控填错了性别，最后无奈之下把性别加粗了。嗯，是加粗了，不是改了。

3. 论文的准备

首先diss一下很多人把收不到NLP算法岗offer归结于没有顶会论文，实际上硕士期间出一篇顶会论文对于绝大部分硕士来说还是非常难的，更何况AAAI这类顶会自带灌水嫌疑，如果面试过程中也给人感觉是水文的话更不算什么加分项了。一般来

说，硕士期间出一篇有意义的顶会最少意味着：

1. 研一下学期之前把所有基础打好（不然别人刷paper的时候你刷教材么）
2. 有很负责且很有经验的导师/高年级博士带（纯靠自己摸索的话基本洗洗睡了）
3. 方向没偏（有些导师做的方向实在难出paper）
4. 数学、coding能力过关（学个numpy、tensorflow要花大半年的话可以转行了）
5. 运气好（不是说观点新，实验效果好，写的也ok就能中的，你要相信神马审稿人都会有的）

一般来说研二那一年的最后能投的顶会是在4月之前的，再往后的会议出来录取结果的时候基本校招提前批进行一大半了，简历能丢的都丢出去了，所以真正可以好好做科研的时间基本只有研一下学期+研二上学期。（本科出顶会的大佬忽略）

而对于大部分人来说，很难一下子就有idea，然后一下子就做出来了好实验效果，一下子论文写的也很漂亮，一下子就中了顶会的，基本都要从水会投起，摸索套路，所以一无所有的情况下非顶会不投几乎等同于作死（当然水某些A类顶会的话另说）。但！是！尽自己的努力去水一篇非顶会还是不难的（比如CCF C类的水会，EI检索之类的）。水完水会之后，想进一步水顶会的小伙伴们可以根据这个ddl时间表规划一下自己的时间

-> <https://aideadlin.es/?sub=NLP,ML>

4. 比赛的准备

除了论文，另一个有吸引力而且性价比更高的活动就是打比赛了。个人英雄主义在大课题、大项目里很难体现，但是在一个比赛中却是有了充足的show的机会。但是注意那些小企业小机构办的几十个队伍参赛的小比赛就不要水了，提交个baseline模型就能水个top3挺没意思的，挂简历上会“很掉身价”的。

除了NLP各大顶会和kaggle的比赛，还可以多多关注各个互联网大厂举办的NLP比赛，比如微软的编程之美挑战赛（去年是问答bot）、百度的机器阅读理解大赛、阿里的天池系列比赛等，亲测在各大厂的面试中很有效（非举办方的互联网公司一般也都会关注友商的比赛的）。

另外，一定一定要打跟目标岗位match的比赛，底线是NLP比赛。比如你想做chatbot，却光打一些数据挖掘类的比赛，哪怕top 5也意义不大，但是这时文本匹配、生成、问答相关的比赛哪怕排名一般（差不多容忍到队伍数的10%）也完全可以强行写上去。已经有名次很好但是match程度不高的比赛了怎么办？简历上一句话带过吧，突出强调和大写特写只会让人觉得你跟岗位不match。

具体到比赛过程中，如果你要做NLP，千万不要靠堆开源模型做ensemble上分，这样虽然能为你争取到面试机会，但是这样基本没有任何创新，只会让面试官觉得你是个优秀的搬砖工程师和调参小能手，如果简历上也没有亮点，很容易被挂（身边有活体例子）。对于面试来说，优秀的单模型超级好用！另外最好把顶会SOTA也拿到比赛数据集上跑一下，这样面试时更有说服力。

5. 实习与项目的准备

一份儿好的实习经历确实可以加分，比如MSRA、百度自然语言处理部、阿里idst、腾讯AI lab这些部门的research经历或者大厂match的业务部门的出色工作都会让面试难度降低不少，然鹅小夕亲测这并不是必须的（心疼的抱抱导师不放实习的自己(。 久。 ㄟ。))。同时，小夕也亲测实验室的大项目一般木有多少吸引力，远不如上面那几个地方的实习经历值钱，所以抱怨自己实验室没项目的童鞋可以stop了，去找实习机会吧。

6. 其他

最后啰嗦一下，顶会paper、top比赛、大厂研究院实习都不是必须的，但是最好它们取或运算后为真。从小夕身边的小伙伴的情况看，只要取或运算后为真，一般BAT的核心研究部门或者核心业务部门最少也能拿一个offer。

如果很不幸成为了“三无人员”，那么一定要保证扎实的数学、NLP、coding能力和最少一个研究方向的专精，能够在面试时表现出超出简历描述的能力，这样也会非常打动面试官的，毕竟谁都想招一个潜力股嘛。

干货篇

1. 面试中的基础知识

这一篇说起来都是泪，小夕曾经努力挤出时间复习，手撸了一遍逻辑回归、最大熵、决策树、朴素贝叶斯、SVM、Adaboost、EM、HMM、CRF的推导，结果发现从来没被问到过。。。从来没被问到过。。。从来没。。。。

然后很奇葩的竟然被问到过TCP三次握手，hadoop中的shuffle机制，linux的find命令怎么用。。。 (from 今日头条)

虽然实现NLP的方法基本离不开机器学习与神经网络，但是如果按照前面简历篇讲的准备简历内容，其实在NLP岗的面试中很少直接考察ML和NN的理论知识。那考察什么呢？当然是考察关键词呀！所以总结一下自己简历的关键词，然后展开复习吧！

下面小夕拿自己举个例子。小夕简历上出现的NLP关键字如

问答、MRC、对话、匹配、词向量、迁移、分类、分词、POS、NER等

下面是面试中考过的基础知识举例

trick: 方向不match的面试官喜欢考察词向量和文本分类相关的知识

模型篇

- SGNS/cBoW、FastText、ELMo等（从词向量引出）
- DSSM、DecAtt、ESIM等（从问答&匹配引出）
- HAN、DPCNN等（从分类引出）
- BiDAF、DrQA、QANet等（从MRC引出）
- CoVe、InferSent等（从迁移引出）
- MM、N-shortest等（从分词引出）
- Bi-LSTM-CRF等（从NER引出）
- LDA等主题模型（从文本表示引出）

训练篇

- point-wise、pair-wise和list-wise（匹配、ranking模型）
- 负采样、NCE
- 层级softmax方法，哈夫曼树的构建
- 不均衡问题的处理
- KL散度与交叉熵loss函数

评价指标篇

- F1-score
- PPL
- MRR、MAP

可以看出，其实面试考察的基础知识的内容跟简历中的关键字高度相关有木有。虽然偶尔也会问到一些超出关键字限制的范围，但是非常非常少，而且基本都是些跟研究方向无关的基础知识。当然，这个前提是按照前面简历篇的建议来认真构造“算法研究”的板块呐。身边不少小伙伴吐槽面试官问的问题很偏，跟自己的方向差别大，然而都木有反思过自己的简历到底有没有告诉面试官自己擅长什么，遇到不match的面试官的话肯定被问成狗哇。（当然，遇到完全不看简历，方向又跟自己不match的面试官的话可以善意的提醒“抱歉我不是做这个方向的”）

2. 面试中的设计/方案题

除了基础知识，有的公司还会出一些开放性的设计题（尤其在最后一两轮面试或者是为SP、SSP设置的加面时），解这些设计题主要还是靠项目和比赛的经验积累，切忌不要只拿论文说事儿，要站在以最小代价来解决问题的角度出方案，而不是非要用上最新的论文。

解这类题的一条基本原则是，能用规则解决就不要用数据，能用简单特征工程解决就不要上大型神经网络，上大型网络时尽量不要使用深度LSTM这类推理复杂度太高的东西。

最后总结一下，精心的准备好简历，那么从巨头到startup的面试其实都大同小异（个别奇葩面试套路的公司除外）。hold住自己的简历和研究方向，白板编程别太差，面试基本都能过。另外，小夕参加的都是提前批，不太清楚正式批会不会有变化。不过貌似正式批的NLP岗少得多，尽量不要把重心放正式批哦

TIPS篇

1. 一定不要错过提前批！一定不要错过提前批！一定不要错过提前批！不要相信错过提前批还有正式批的鬼话，很多核心部门的NLP岗的hc在提前批就用光了！
2. 提前批不要拖到末尾！尤其百度自然语言处理部的坑，面试当天就给口头offer，先占先得！
3. 阿里基本是远程面试，6轮面试都没编程题！简历内容准备好就可以直接投阿里了！
4. 内推之前千万不要把简历挂到腾讯校招系统，填上意向部门都没用，不match的其他部门很可能以迅雷不及掩耳之势强行捞起你的简历，然后你的腾讯之旅就举步维艰了（我跟另外俩小伙伴血的教训。。。）
5. 如果只是想找算法岗，但没有研究方向，没刷几篇paper，不过研发能力强，基础编程题解的快，工具用的熟，那就去投头条吧。（亲测它提前批的三轮技术面都没学术味儿，所以优势在对立面的小伙伴谨慎）
6. 网易游戏可能是国内唯一一家不用在性别栏2选1的企业（测评界面），无偿点个赞（其他大厂能不能效仿一下啊喂，优秀的跨性别小姐姐小哥哥那么多，照顾一下嘛\\(//▽//)\\)
7. 谨慎投递跟第三方招聘网站合作的企业，除非你想换手机号了\\(//▽//)\\



文章已于修改