



机器学习算法
与自然语言处理



中国中文信息学会青年工作委员会
Chinese Information Processing Society of China, Youth Working Committee

算法类岗位求职经验分享

刘乾

qian.liu@buaa.edu.cn

北航与微软亚洲研究院联合培养博士

The 3rd Seminar @ MLNLP

目 录

- 01 投递篇：挑选岗位
- 02 简历篇：准备简历
- 03 面试篇：代码基础
- 04 面试篇：项目展示
- 05 面试篇：行为面试
- 06 彩蛋篇：我的秋招



投递篇: 公司类型

国内外企



Microsoft



Amazon



Shoppe



Apple

...

国内大厂



腾讯



阿里巴巴



字节跳动



百度

...

新型机构



上海人工智能实验室
Shanghai Artificial Intelligence Laboratory



智源人工智能研究院

...

投递篇: 岗位类型



科学研究: 自由



以高影响力论文为目标



需要强自我驱动与探索



无任何落地要求, 适中



业务场景: 核心



研发算法增长业务场景



业务场景可能变化频繁



与业绩挂钩, 波动大



研究落地: 成长



同时做研究与系统原型



研究方向与合作方挂钩



有落地要求, 适中

投递篇：挑选岗位

1. 每种类型的岗位建议至少选择一个典型代表参与面试，不要在面试前轻易否定任何一个岗位，等到最终的offer之后再做抉择。
2. 偏研究和偏业务没有好坏，只是不同的职业规划，并且它们都需要较为繁重的脑力劳动。不同岗位最核心的需求不同，把握好优先级。
3. 7月份是提前批，9月份是正式批，12月份大厂集中秋招结束。早做准备，越早投递简历越好，不要总等「万事俱备」。

简历篇: 好的模板

使用 Latex 构建简历，内容保持在两页内。

- 突出论文，推荐模板 <https://github.com/billryan/resume>
- 突出实习，推荐模板 <https://github.com/posquit0/Awesome-CV>

- Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, Dongmei Zhang, **You Impress Me: Dialogue Generation via Mutual Persona Perception**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL-2020)*
- Qian Liu, Bei Chen, Haoyan Liu, Lei Fang, Jian-Guang Lou, Bin Zhou, Dongmei Zhang, **A Split-and-Recombining Approach for Follow-up Query Analysis**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-2019)*

- Haoyan Liu, Lei Fang, Qian Liu, Bei Chen, Jian-Guang Lou, Zhoujun Li, **Leveraging Adjective-noun Phrasing Knowledge for Comparison Relation Prediction in Text-to-SQL**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-2019)*
- Qian Liu, Bei Chen, Jian-Guang Lou, Ge Jin, Dongmei Zhang, **FANDA: A Novel Approach to Perform Follow-up Query Analysis**. In *33th AAAI Conference on Artificial Intelligence (AAAI-2019)*

SKILLS

- Programming Languages: Python > C# > JavaScript > HTML/CSS = C++
- Deep Learning Tools: FairSeq > AllenNLP = PyTorch > Transformers > Tensorflow

HONORS & AWARDS

Ph.D. Academic Excellence Foundation	2021
Baidu Scholarship Top 20 Worldwide	2020
Graduate Student National Scholarship	2019, 2021
Beijing Outstanding Graduate Awards	2017

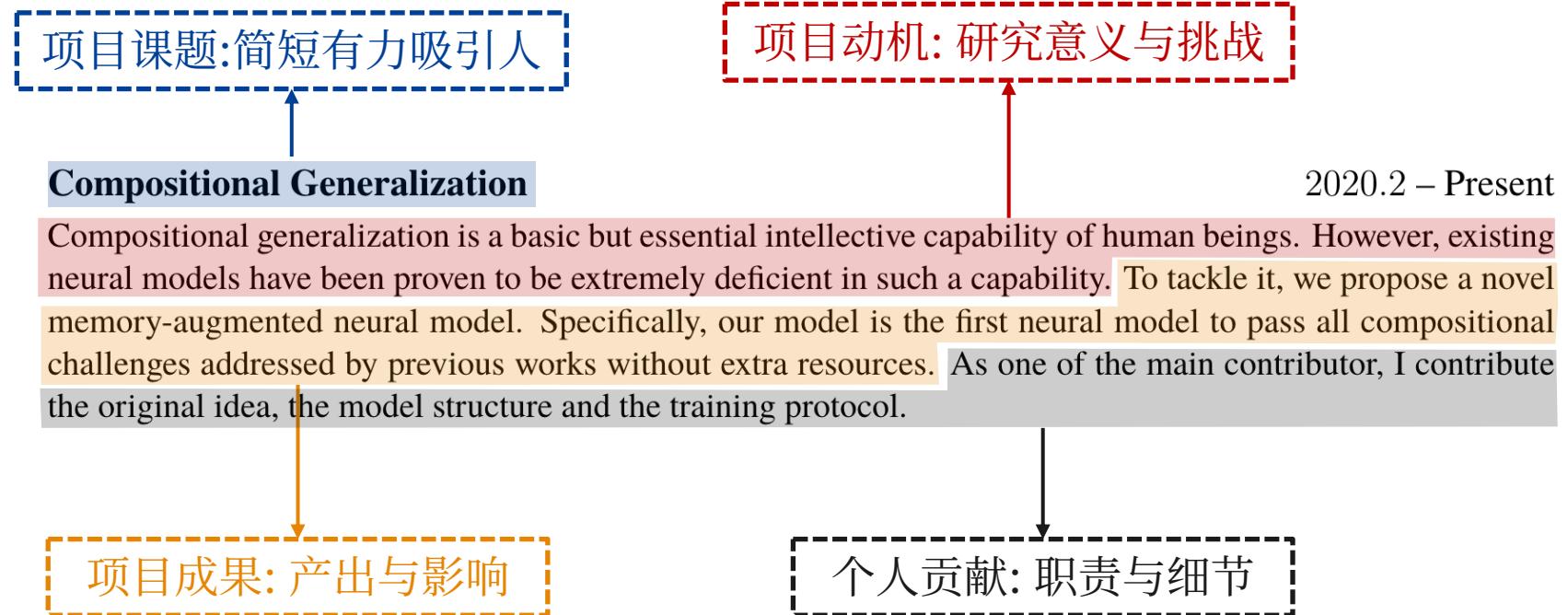
The screenshot shows a two-page resume template for Byungjin Park. The top page features a header with the name 'Byungjin Park' in large serif font, followed by 'SOFTWARE ARCHITECT • SECURITY EXPERT'. It includes contact information like phone number, email, website, and social media handles. Below the header is a quote: "Be the change that you want to see in the world." The second page is titled 'Summary' and describes Park's current role as a Site Reliability Engineer at Kasa, mentioning his experience in backend development, automation, and security. It also highlights his interest in problem-solving and learning new technologies. The 'Work Experience' section lists a position at Omnious. Co., Ltd. in Seoul, S.Korea, from Jun 2017 to May 2020. The responsibilities include managing hybrid infrastructure, automating CI/CD pipelines, designing microservices architecture, and implementing serverless functions. The resume uses a clean, modern design with a white background and black text.

简历篇: 主要内容

简历内容主要包括:

-  基本信息 (姓名, 邮箱, 电话, 个人主页)
-  教育经历 (本科成绩 Top N / M %, 研究生实验室, 研究生导师)
-  实习经历 (实习性质, 实习部门)
-  项目经历 (项目课题, 时间, 项目动机、项目成果与个人贡献)
-  发表论文 (论文名, 作者, 会议/期刊, Oral/Spotlight)
-  获奖经历 (校级及以上奖项, 或比较特别的荣誉)

简历篇: 项目经历



简历篇：准备简历

- 面试时往往会根据项目经历来提问，尤其是第一段项目经历。我们要做到对项目经历的每一个字都非常熟悉。



这简历真不错 我要了

- 简历整体要保持简约清爽，避免使用花里胡哨的颜色，叙述最重要的部分，搭配一些简单的图标即可😎。

面试篇：算法基础

算法基础主要包括如下四方面， 表示我个人认为的重要程度：

- 计算机学科基础(编程语言特性、计算机课程)
- 机器学习基础(SVM、逻辑回归、EM、反向传播、集成)
- 领域相关基础(CV/NLP/预训练模型架构)
- 应用领域基础解决新场景(往往是业务相关的综合场景)

面试篇: 领域相关基础

- Attention 公式
- Transformer 实现 (推荐 <https://github.com/karpathy/minGPT>)
- Normalization 的原理与好处
- BERT
- 项目经历涉及的基础知识

面试篇: 手撕代码

从手撕代码的难度和重视程度来排个序 (仅根据个人秋招经验)



上海人工智能实验室
Shanghai Artificial Intelligence Laboratory



智源人工智能研究院



面试篇：手撕代码

- 系统地预复习算法 (推荐公众号 [labuladong的算法小抄](#))
- 熟悉面试时的代码环境 (推荐 [leetcode](#), [牛客网](#))
- 背一套好用的算法模板 (推荐 [九章算法](#) 的模板)
- 抱佛脚刷企业高频题 (推荐 [codetop.cc](#))
- 养成好的代码习惯，注意注释、变量名、函数拆解

面试篇: 手撕代码



- 二分查找 ([leetcode 33. 搜索旋转排序数组](#))
- 排序算法 ([leetcode 912. 排序数组](#))
- 快速选择 ([leetcode 257. 数组中的第K大个数](#))
- 二叉树相关 ([leetcode 236. 二叉树的最近公共祖先](#))
- 深度/广度优先搜索 ([leetcode 200. 岛屿数量](#))
- 动态规划 ([leetcode 42. 接雨水](#))

面试篇：项目展示

两种展示形式

1. 时长约20min，包括  自我介绍， 获奖经历， 最喜欢的项目。
2. 时长约40min，除上述外，还要包括  研究框架， 未来工作。
3. 展示①适用于所有算法岗，展示②尤其适用于需要讲Talk的面试。

面试篇: 项目展示



标题页

Towards Data-Efficient Semantic Parsing

Qian Liu (刘乾)
qian.liu@buaa.edu.cn
Joint Ph.D. of BUAA and MSRA

Sea AI Lab Candidate Interview



自我介绍

Introduction

YOUNG

Qian Liu Joint Ph.D. of BUAA and MSRA

- Education Background**
 - B.S. and Ph.D. in Beihang University, Computer Science
 - Ph.D. Supervisor: Qipeng Zhao & Jian-Guang Lou
 - Internship at MSRA from July 2016 to Now
- Research Interest**
 - Natural Language Processing
 - Semantic Parsing



获奖经历

Achievement

Research	15	Top Conf. Papers
	NeurIPS	
	AAAI	
	ACL	

Industry	3	Microsoft Products
	PowerApps	
	Excel	
	Power BI	

Honor	20	Baidu Scholarship
	Stanford	
	CMU	
	Tsinghua	



未来工作

Direction 3: Intelligent Programming

What is id of the car with the max capacity?
Sorry, I do not understand "capacity". What does it mean:
A. Horsepower B. MPG C. Make D. None
Choose A.
SELECT Id FROM CARS_DATA ORDER BY Horsepower DESC LIMIT 1

Generalizable
Understand user questions across distribution, source and domains.

Interactive
Interact with users when it does not understand the question.

Trustworthy
Align with user feedback through interpretable design.



项目介绍

Part 2. Unleash the Power of Task Data

Previous Work
SELECT Weight FROM CAR_B_DATA ORDER BY Horsepower DESC LIMIT 1
SQL Executor → 3693

Our Work
SQL Executor → 3693

Latent Program Learning
Learning latent program unleashes the power of task data, even making it compete with program data.

Combine LMs with rules makes it possible to proxy real data with synthetic data without crowd sourcing.

Liu Q et al., TiAPEX: Table Pre-training via Learning a Neural SQL Executor. ICLR 2022 Top Score (1/31, Nov. 9).



研究框架

Overview of Talk

Goal

- Part 3. Proxy Real Data with Synthetic Data
- Part 2. Unleash the Power of Task Data
- Part 1. Reduce the Size of Program Data

Achievement

- Combine LMs with Rules XYZ% acc in 1 week
- Latent Program Learning 4% than SOTA across 4 datasets
- Memorize to Generalize 100% acc on known testbeds

面试篇：展示原则

- 不要指望一次功成，展示效果是随着迭代次数逐渐变好的。可以找身边人给出反馈，也可以在与面试官的互动中学习。

 刘乾-阿里星级委员会面试-1.pptx	9/2/2021 11:09 PM	Microsoft PowerP...	43,072 KB
 刘乾-阿里星级委员会面试-2.pptx	9/10/2021 7:16 PM	Microsoft PowerP...	21,333 KB
 刘乾-阿里星级委员会面试-3.pptx	9/13/2021 2:21 PM	Microsoft PowerP...	15,718 KB
 刘乾-阿里星级委员会面试-4.pptx	9/14/2021 7:52 AM	Microsoft PowerP...	15,294 KB
 刘乾-阿里星级委员会面试-5.pptx	9/14/2021 7:56 AM	Microsoft PowerP...	20,077 KB

- 项目展示选择简约的模板 (推荐 [officeplus](#))，注意**颜色搭配**，少用字 (推荐 [思源, Lato](#))，多用图标 (推荐 [emoji, iconfont](#))，不要讲太多公式！

面试篇: 项目介绍

背景与挑战(1/3)

整个展示中最重要的部分，首先要让面试官懂，其次要引起他的兴趣。

Previous Approaches

No model can successfully solve compositional challenges on SCAN!

Model	Add Jump	Around Right	Length Generalization
Seq2Seq(NYU)	1.2	2.5	13.8
CNN(Facebook)	69.2	56.7	0.0
Syntactic Attention (MILA)	91.0	28.9	15.2
CGP (Baidu)	98.8	83.2	20.3
GCA (MIT)	86.0	82.0	-
Meta Seq2Seq (Facebook)	99.9	99.9	16.6
Equivariant Seq2Seq (Cambridge)	99.1	92.0	15.9

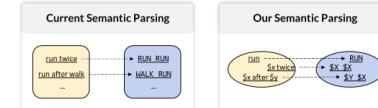
#17

动机与方法 (1/3)

动机部分要讲清楚。
方法部分不要讲太细节，讲清楚方法的框架即可。

Abstraction ↗ Compositional Generalization

Natural language and programming language are both **abstractive**, of the sort that can be captured by **variables with symbolic functions**.



#19

影响与启示 (1/3)

讲出的地方，并把方法引申的规律陈列出来，震撼面试官。

Insight: Decomposition for Better Generalization

Performing task **decomposition** (i.e., decomposing a complex task into relatively simple tasks) improves the **compositional generalization**.



#25

面试篇：行为面试

要提前了解公司的**文化观与价值观**，寻找与自身的契合点来回答问题或提问。

1. 请你先做个简单的自我介绍？

- **开场礼貌：**面试官好，非常荣幸有这样一个机会跟您聊一下。
- **基本信息：**我叫刘乾，目前是北京航空航天大学计算机学院直博五年级。
- **学习经历：**我的导师是虚拟国家重点实验室的赵沁平老师。在过去的五年中，我也有幸参与到微软亚洲研究院联合培养博士的项目中，指导老师是楼建光首席研究员。
- **研究方向：**在两位导师的指导下，我的主要研究方向是自然语言处理，主要研究兴趣是语义解析。
- **获奖经历：**围绕着数据高效的语义解析，我目前已经 在国际顶级会议发表相关论文 15 篇，也因此荣幸获得了 2020 年百度奖学金提名等荣誉。

面试篇：行为面试

要提前了解公司的**文化观与价值观**，寻找与自身的契合点来回答问题或提问。

2. 你对未来的职业规划是什么，为什么选择当前这个岗位？

➤ 抓住岗位的核心特点，要了解且突出目标公司/岗位的独特之处。

3. 我注意到你有在 ABC 实习过，为什么不考虑转正留在那里？

➤ 讲出对实习岗位的顾虑，并比较两个岗位对自己而言的区别。

面试篇：行为面试

要提前了解公司的**文化观与价值观**，寻找与自身的契合点来回答问题或提问。

4. 你的同事一般会怎么评价你？

➤ 这时候要自信地讲自己的优秀品质，并高亮与公司文化对应的部分。

5. 你对我们这里还有什么问题吗？

➤ **原则**：视面试官级别提问，可以是很难回答的问题，但注意要保持礼貌！
➤ **部门同事**：具体的问题，比如同事的研究方向，对自己面试的反馈等
➤ **部门领导**：小组的问题，比如小组内的方向，研究与工程的占比等
➤ **公司高层**：宏观的问题，比如整个大部门的发展战略

彩蛋篇：我的秋招投递时间线



机器智能

2021.07.09

三轮Talk面+一轮笔试

阿里星面试

2021.09.15



NLP

2021.08.10

三轮技术面

AIDU面试

2021.08.27



Tencent
AI Lab

NLP

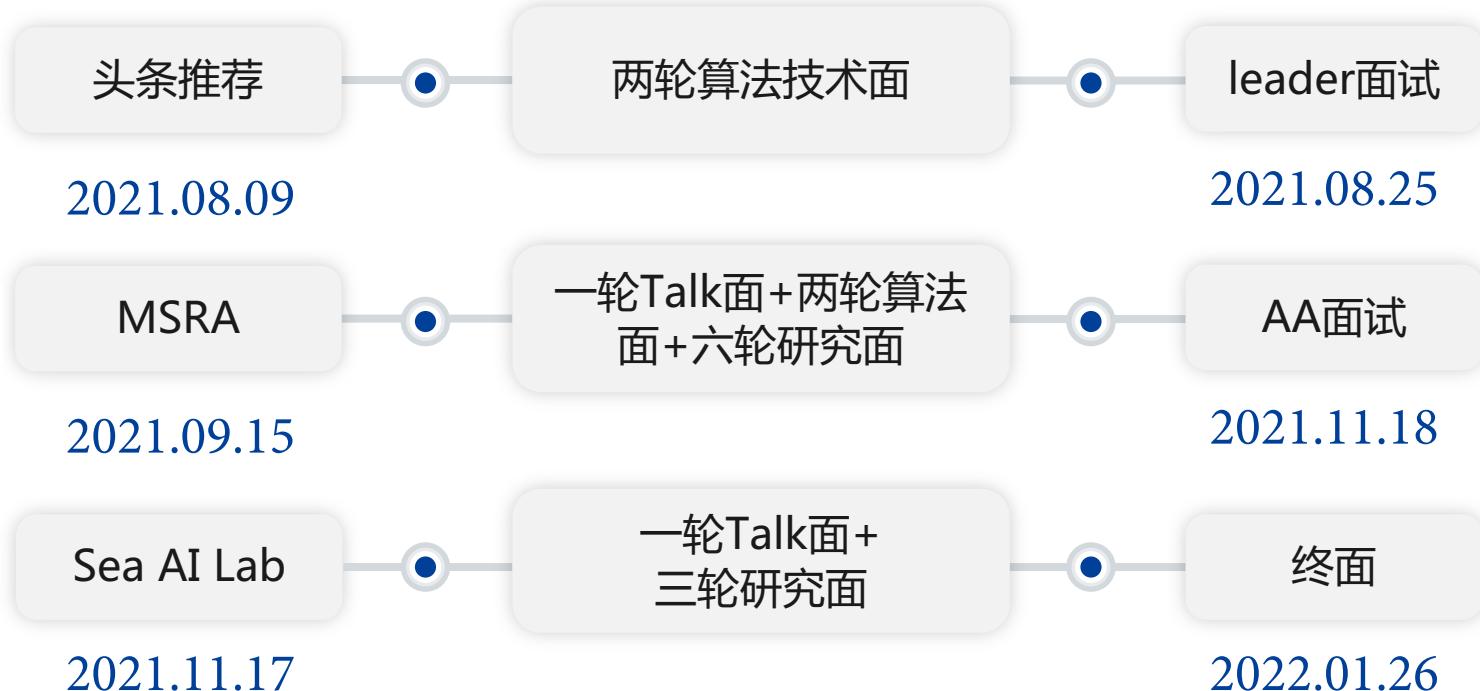
2021.08.17

三轮技术面+一轮总监

大咖面试

2021.10.20

彩蛋篇：我的秋招投递时间线

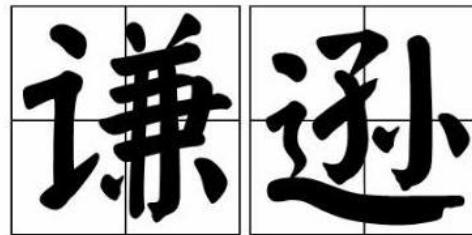


彩蛋篇：我的秋招关键字

打听

要充分发挥主观能动性减少信息差，多找师兄、好友、心仪公司的同事打听。

彩蛋篇：我的秋招关键字



不管与谁 (内推同事, 部门leader, 招聘HR) 进行沟通, 都要保持一个良好的态度。

彩蛋篇：我的秋招关键字



天下没有完美无缺的offer，如果有，那一定是它的缺点还没有被发现。

彩蛋篇：面试季我最感谢的人



李永彬

阿里巴巴达摩院 Conversation AI 团队负责人

招聘KBQA / TableQA等智能问答方向的研究实习生与员工，联系 binyuan.hby@alibaba-inc.com。

褚智慧

上海人工智能实验室招聘总监

面向全球招募顶尖AI科研人才，包括但不限于人工智能基础理论，核心技术，开放平台等，联系左侧微信。

楼建光

微软亚洲研究院DKI组首席研究员

招聘语义解析/代码生成方向的研究实习生，联系 bei.chen@microsoft.com

Thanks & QA

刘乾

qian.liu@buaa.edu.cn

北航与微软亚洲研究院联合培养博士