

# 唐宇帆

籍贯：浙江省杭州市 | [tangyf0522@gmail.com](mailto:tangyf0522@gmail.com) | (+86)156-5188-5707  
NLP 算法实习生



## 教育经历

代尔夫特理工大学大学  
计算机科学硕士

荷兰代尔夫特市  
2022.09 – 2024.08

南京航空航天大学  
计算机科学与技术学士

江苏省南京市  
2018.09 – 2022.06

- 绩点: 90/100, 排名: 10/132, top 8%

爱丁堡大学  
公派交换生

英国爱丁堡市  
2020.09 – 2020.12

## 工作实习经历

阿里云计算有限公司  
Java 实习工程师

浙江省杭州市  
2021.07 – 2021.09

- 作为项目独立负责人之一开发监控告警管理平台 Xdata, 向业务的管理者以及应用的负责人展示业务与应用健康性能大盘。在这一过程中主要负责开发前端数据与数据库交互过程中会调用到的相应的增删改查功能接口, 同时保证查询性能的稳定性。
- 参与公司相关的销售生态管理业务, 为了增加销售人员的公司政策参与体感, 保证政策的落实度, 参与开发面向销售团队的激励移动端软件。在这一过程中主要负责相应的数据库复杂查询接口的撰写与测试调优。

## 科研经历

Text-to-SQL 研究与实现

江苏省南京市  
2021.12 – 至今

- 基于 seq2seq 的机器翻译模型 mBART、mT5 等模型, 把 Text-to-SQL 任务简化为机器翻译的端到端任务, 利用 transformers 以及 fairseq 等开源工具进行模型的训练和调参。
- 基于 RYANSQL 进行模型改进, 将原先应用于英文 Text-to-SQL 数据集的模型推广到中文数据集上进行实验测试, 对中文 Text-to-SQL 数据集设计出了一套分库的方法以及预处理数据集的脚本。
- 利用模版匹配 GAZP 和机器翻译的手段, 应用现有开源的 DuSQL、CHASE 等中文 Text-to-SQL 数据集训练了 SQL-to-Text 的模型, 在现有数据集上进行数据增广, 扩充现有数据集, 提升模型的泛化能力。

基于 GAN 数据增强的 Bert 模型在新闻文本分类上的研究

江苏省南京市  
2021.09 – 2022.04

- 主要考虑通过 GAN 在对抗训练过程中会生成额外的数据样本, 通过生成的样本作为数据扩充加入到原始数据集中实现数据增强, 进一步提升最终的分类模型的泛化能力, 选用的 GAN 模型为多标签对抗生成网络 catGAN。

## 奖项荣誉

2020-2021, 2019-2020, 2018-2019 南京航空航天大学学业奖学金 (top20%)  
2020-2021 南京航空航天大学优秀学生奖学金 (top15%)  
2019-2020, 2018-2019 南京航空航天大学优秀学生奖学金 (top15%)  
2020-2021, 2019-2020 南京航空航天大学三好学生 (top15%)  
2020 全国大学生数学建模竞赛江苏省赛区  
2021 华为智能基座奖学金

一等奖  
二等奖  
三等奖  
  
二等奖  
一等奖

## 技能/兴趣爱好

- 编程语言: Java, Python, C++, SQL, Matlab