

# 白泽珩

手机：(+86) 13021191066 · 邮箱：zeheng@hgc.jp

性别：男 · 籍贯：北京



## 教育背景

- |   |       |       |                   |
|---|-------|-------|-------------------|
| 东京大学  | 计算机科学 | 硕士    | 2021.04 - 2023.04 |
| • 通过自己自学计算机专业的全部课程通过了东京大学计算机研究生院的招生考试   |       |       |                   |
| 东京大学  | 生物信息学 | 学生研究员 | 2020.10 - 2021.04 |
| • 于 2020 年 10 月起将在线上与东京大学医科学研究所井元清哉教授一同进行科研   |       |       |                   |
| 中国人民大学  | 统计学   | 本科    | 2016.09 - 2020.07 |
| • 主修课程：高等代数、数学分析、微观经济学、统计学概论、概率论、数理统计学、C 程序设计、MySQL 数据库、Excel VBA、非参数统计、随机过程、实变函数、回归分析、时间序列分析、多元统计分析、统计计算、生物统计、空间数据分析 |       |       |                   |

## 校内实践

- |   |               |                   |
|---|---------------|-------------------|
| 中国人民大学统计学院  | 学生会、青年志愿者协会干部 | 2016.10 - 2017.09 |
| • 负责学生会外联部、青年志愿者协会行动部的主要工作，每一项任务都认真完成、尽职尽责，赢得了老师和同学的信赖  |               |                   |
| 中国人民大学统计学院  | 班长            | 2017.09 - 2018.09 |
| 中国人民大学第二十届“创新杯”   | 小组成员          | 2017.09 - 2018.10 |
| • 与同学合作完成调研“农村家庭教育投入激励因素及模式研究”，成功获得学校第二十届“创新杯”三等奖，并获得学生课外学术科技作品竞赛推荐资格，成功入围中国人民大学第二届本科生学术论坛主论坛 |               |                   |
| 中国人民大学“街巷中国”寒假实践项目  | 调查员           | 2018.01 - 2018.04 |
| • 利用寒假时间在苹果园街道进行调研项目，主要内容是向街道办事处了解街道发展情况，以及街道人民的政治意识及生活水平，调研十分成功并顺利结项                         |               |                   |
| 中国人民大学“大学生创新实验计划”   | 小组成员          | 2018.05 - 2019.05 |
| • 与同学合作报名中国人民大学“大学生创新实验计划”中的大学生创业训练计划，“喵嗷屋猫咖啡有限责任公司”创业训练项目作为国家级项目成功结项                         |               |                   |

## 科研经历

- |   |        |                   |
|---|--------|-------------------|
| 探究 VIBRANT 的泛化能力及改进   | 个人研究课题 | 2020.10 - 至今      |
| • 指导教师：东京大学医科学研究所教授 井元清哉、副教授 张耀中  |        |                   |
| • 阅读相关文献，利用 Ubuntu 和 Docker 对研究相关的工具，如 VirSorter, VirFinder, MARVEL, VIBRANT 等基于数据库的模型、机器学习模型、深度学习模型深入掌握。 |        |                   |
| • 在 NCBI 收集相关基因序列数据并进行标准化处理   |        |                   |
| • 在超级计算机上运行深度学习模型 VIBRANT   |        |                   |
| • 对 VIBRANT 提出改进，如使用 XGBoost 代替神经网络或使用 XLNet 代替隐马尔可夫模型  |        |                   |
| 台湾空气污染及流行病相关关系  | 小组成员   | 2018.11 - 2020.07 |
| • 指导教师：中国人民大学统计学院副院长 李扬教授、辅仁大学副校长 谢邦昌教授、中国人民大学统计学院讲师 王菲菲教授  |        |                   |
| • 我们使用 Python 通过网络爬虫获取 2010 年至 2018 年台湾气候的每小时数据，并使用 R 将相同监测站的数据合并   |        |                   |
| • 利用 K-means 对异常值进行补全，利用 Kriging 对数据进行整理和分析，并做相关热力图和散点图的数据可视化   |        |                   |
| • 利用空间自回归、变系数回归等手段进行分析，并已有研究成果  |        |                   |

- 同时我在组内会带领新加入的成员熟悉数据和代码，并教会他们如何高效地使用 R。迄今为止，小组内相关的可视化 R 代码全由本人所写，并且相比之前的版本大大降低了复杂度

## 2019 世中联中医药大数据白皮书 编者 2019.07 - 2019.12

- 指导教师：北京大学公共卫生学院生物统计系主任，北京大数据研究院中医大数据中心主任 周晓华教授、北京大数据研究院中医大数据中心助理主任 黄新霆博士
- 我们利用爬虫爬取近几年开展有关中医药大数据的企业、高校、医院、政府等信息
- 通过电子邮件联系的方式，对这些数据进行校正和完善
- 对这些数据进行可视化，使得报告更加易懂和直观
- 本人完成了该白皮书占据主要篇幅的：《2019 世中联中医药大数据白皮书：企业篇》

## 主动健康和人口老龄化（北京“双一流”项目） 小组成员 2019.06 - 2019.07

- 指导教师：中国人民大学统计学院院长 王晓军教授、中国人民大学统计学院讲师 王菲菲教授
- 我们利用分布式爬虫算法爬取近年人们对于主动健康和老龄化的舆情文本
- 基于 LDA 主题模型优化并建立适合本题目的情感分析模型

## 实习经历

---

### 美团 数据产品实习生 2020.10 - 至今

- 对数据产品涉及指标进行规范化、标准化治理
- 撰写日报《使用数据表报备流程规范》
- 与研发部门合作向集团领导提供报表
- 协助研发对集团内部数据产品进行迭代

### 小米集团 数据分析实习生 2019.11 - 2020.02

- 负责小米海外应用商店（GetApps）的数据管理与分析
- 利用 Hive 对这些数据进行存储和提取，并对可以增加收入的方法进行寻找和报告
- 其中的部分数据打点定义做出了完善和修正，并尝试用 LSTM 对日活数据进行分析和预测

### 中铁建资产管理有限公司 软件开发部实习生 2019.01 - 2019.02

- 对即将上线的 APP “铁建金服”进行初步的熟悉
- 了解这个 APP 的代码内容，并协助开发人员进行测试

## 境外交流

---

### 加州大学伯克利分校 暑期学校课程 2018.07 - 2018.08

- 2018 年暑期成功进入美国加州大学伯克利分校（University of California, Berkeley）学习暑期学校课程（生物统计学及脑科学），并获得平均成绩 97 以上的优异成绩

## 职业技能

---

- 语言：中文、英语（六级：540，托福：95）、日语
- 操作系统：Windows、Linux
- 编程语言：R、C、Python、Visual Basic、SparkSQL、HQL
- 专业软件：SPSS、Stata、Office、Acrobat、Xmind、LaTeX、Notepad++
- 开发者工具：Docker、Singularity、VS Code、Xshell、Xftp、Visual Studio、PyCharm、Rstudio、Ubuntu