

# 温慧安

性别：男                      出生：92.10.24  
民族：汉                      籍贯：江西赣州

期望岗位：  
数据分析岗

130-9733-1225  
526166357@qq.com

## 专业技能

- 编程语言：Python、Matlab、MySQL、C语言（全国计算机二级）；
- 数据分析/可视化：PowerBI、Tableau、Python、Matlab、Excel；
- 办公软件：Word、PPT、Excel、Visio、思维导图；
- 系统环境：Linux 及其常见指令；
- 算法模型：了解机器学习常用模型如 LR，决策树，GBDT，RF，Bagging等；
- 专业背景：良好通信与计算机技术背景，掌握计算机网络、数字信号处理及概率论与统计等；
- 英文水平：良好英文文献检索和阅读写作能力（CET6）。

## 教育背景

2014.09 – 2017.06	重庆邮电大学	硕士	电子与通信工程
2010.09 – 2014.06	南昌航空大学	学士	通信工程

## 工作经历

2018.06 – 至今	上海南燕保险科技（C轮，SIG创投）	数据分析师
<ul style="list-style-type: none"><li>➤ 中台及数据平台需求分析；</li><li>➤ 高端医疗企补理赔金额预测模型；</li><li>➤ 数据库与 Salesforce 数据 ETL 逻辑及流程梳理与优化；</li><li>➤ 负责日常 ERP、理赔数据清洗、挖掘与分析，定期形成数据报告；</li><li>➤ 外部 web 或 app 爬虫开发，外部数据爬取及分析；</li><li>➤ 内部运营数据可视化及分析，实时监控市场部 KPI 完成情况。</li></ul>		
2017.07 – 2017.11	中国电信惠州分公司	云计算网络维护
<ul style="list-style-type: none"><li>➤ 市政务云网络例行维护，定期进行政务云主机漏洞扫描，撰写相关安全报告。</li></ul>		
2014.09 – 2015.09	重庆电力科学研究院	计量部实习生
<ul style="list-style-type: none"><li>➤ 曾参与国家电网横向子项目的项目申报书撰写以及资料调研工作；</li><li>➤ 负责产品性能测试、有关数据测试与分析、后期仿真研究，项目总结工作，以及相关专利论文撰写。</li></ul>		

## 项目经验

2019.10 – 至今	中台及流量数据产品	南燕保险科技
<p>中台需求调研、分析，功能优化；</p> <p>神策数据埋点需求分析、事件设计、项目上线；</p> <p>数据平台需求分析及推动项目上线。</p> <ul style="list-style-type: none"><li>◆ 中台分销模块核心需求分析及设计，参与表结构设计，协同 IT 优化平台功能；</li><li>◆ 结合神策 SDK，利用事件模型获取代理人及 C 客行为数据；</li><li>◆ 利用代理人行为、业务数据构建数据平台，依托 AAARR 海盗模型提供代理人数据产品服务，涵盖指标包括 UV、PV、保费、转化率等。</li></ul>		
2020.10 – 2021.03	高端医疗理赔金额预测	南燕保险科技
<p>机器辅助人工高端医疗个/团单核保定价环节，及时把控风险点。</p> <p>数据提取、清洗、整理；</p> <p>结合核保助手（现有人工经验）与机器学习方法预测结果综合评估字段选取；</p>		

样本处理、模型框架及算法选型，综合效果评估。

- ◆ 数据库表间关系及字段梳理，使用 Oracle SQL 完成数据提取；
- ◆ pandas 库完成数据清洗、整理，采用 sklearn 库 RF 模型完成特征选取；
- ◆ 为考虑高理赔额出险情况，将不同 coverage 理赔金额进行分层，框架选用先分类再回归，基于贝叶斯完成不同层次理赔下的理赔金额加权和作为理赔金额预测；
- ◆ 使用 sklearn 库的 XGBoost 分类模型得到类别概率，综合 RF 回归模型完成预测；
- ◆ Tableau 可视化分析及评估。

2018.10 - 至今

日常数据分析及整合

南燕保险科技

核保：快速对接保司智能核保问卷，处理问卷数据为标准 JSON 格式；

续保：关联线下续保保单，找出到期未续保单，挽回潜在流失保费；

理赔：各保司返回理赔数据清洗、去重、整理入库；

新老后台过渡：中台整合优化，新老系统数据清洗、整理、同步。

- ◆ 使用 pandas、re 等模块快速构建问卷问题树，构建开发所需 JSON 结构；
- ◆ 预定义续保客户规则，利用保单、客户、产品数据关联保单；
- ◆ 各保险公司理赔数据清洗、转换、规整，并加载至 Salesforce 并进行损率分析；
- ◆ 使用 kettle ETL 工具，对新老系统数据进行抽取、转换、加载，保证新老系统同步稳定运行。

2019.10 - 2020.01

外部同业/异业企业客户获取

南燕保险科技

B端客户：通过天眼查、招聘网站、保监会等多个数据源，获取企业信息，提供业务外呼洽谈合作；

医院：通过微医、好大夫等获取医院信息，提供 TPA 业务员与医院展开合作；

微信无资质同业：利用搜狗微信，通过保险类关键字获取同业微信用户，提供业务展开合作

- ◆ 验证码识别，采用 tesseract 模块，对验证码图片进行二值化、裁剪后处理识别；
- ◆ 通过 Fiddle 抓包或 XMLHttpRequest，找到关键接口，进行相关请求；
- ◆ Cookies 池、UA 池、Token 池、Proxy 池反反爬，高效获取数据；
- ◆ 通过 mysqlclient 连接 MySQL 库持久化获取数据；
- ◆ 结合 Salesforce 构建管理平台，业务员通过平台获取相关信息。

2018.10 - 至今

可视化业绩大屏

南燕保险科技

为各团队完成年度指标，对接内部 Salesforce，构建可视化仪表盘，及时了解各团队业绩完成情况。

- ◆ 选择可视化工具 PowerBI 同时与 SalesForce CRM 数据库建立连接；
- ◆ 根据需求，构建多张数据表及可视化图，如折线图、柱状图等；
- ◆ 数据清洗规整，同化异构数据；
- ◆ 结合微软 PowerBI Embed 嵌入式开发，研发 web Dashborad，以供移动端查看；
- ◆ 通过智能TV 浏览器登录网站、显示，并及时更新。

2018.09 - 2019.03

ERP 数据分析

南燕保险科技

针对 ERP 系统中数据，进行可视化分析，了解近期各经代机构保费规模及保单情况，主要涵盖寿险、意外险、财险、健康险为主，了解市场畅销主流保险产品及其竞品，以期找到影响保险销售的主要因素。

- ◆ 通过 python pandas, numpy 库完成数据清洗与规整；
- ◆ 通过 Tableau 可视化工具完成多维度数据可视化；
- ◆ 主要图表包括，柱状图、折线图、热度图及词云图等；
- ◆ 数据分析并形成 PPT 报告。

2018.09 - 2018.10

保险类微信公众号文章爬虫

南燕保险科技

爬取主流保险类微信公众号文章，并通过文章进行简单文本分析，了解用户或代理人主要关注热点。

- ◆ 通过虚拟安卓模拟器，结合 **Fiddle JS** 脚本，获取公众号历史文章链接至本地；
- ◆ 通过 **Fiddler** 抓包，获取 HTTP 协议中的 POST 或 GET 数据，通过 **request** 库 对构建数据包，完成文章链接库的爬取；
- ◆ 通过 **LSI**、**LDA**、**NNMF** 对文本生成主题模型，同时使用 **word2vec** 对文本进行分析。

#### 2018.06 - 2018.08 蜂系推荐小助手

南燕保险科技

通过 CRM 中已有样本数据，分析用户购买情况，从用户不同特征，利用不同模型如 LR、XGBOOST、RF 等，构建推荐模型，找出较优特征。最后通过微服务 **Flask** 库部署上线。

- ◆ python pandas 库完成数据清洗，通过 python sklearn 库 k-means 聚类算法修正部分出错数据；
- ◆ 模型训练及评估，包括交叉验证、P-R曲线图、F1值；
- ◆ 特征重要度分析，通过热度图，柱状图等可视化特征重要度；
- ◆ 选定模型，Linux 上部署 **Flask** 微服务上线。

#### 2016.09 - 2017.03 基于多节点的低压电力线载波通信信道建模研究

重庆邮电大学

该项目为国网重庆电科院合作项目，通过传输线理论，结合最小二乘估计与图论相关算法，完成对多节点电力网络下具有一定相关性的信道建模；通过Class-A噪声概率密度函数，构建随机噪声生成模型。

- ◆ 通过有条件高斯分布，生成泊松随机数，由泊松随机数产生处在不同状态（方差）下的高斯噪声分布随机数，进而得到Class-A随机脉冲噪声；
- ◆ 通过Matlab曲线拟合工具箱对电力传输线特征参数作最小二乘估计；
- ◆ 基于双导体传输线理论，通过最短路径Floyd算法以及基于广度优先的搜索算法，在给定的低压电力网络拓扑下，生成具有整体网络特性的电力线信道响应；
- ◆ 基于System Generator平台在FPGA上对该建模方法进行实现，为载波模块测试提供测试平台。

针对该项目完成硕士毕业论文撰写，获评重庆邮电大学“2017年优秀硕士学位论文”，同时发表相关中文核心论文1篇，相关专利受理2篇。

#### 2016.03 - 2016.08 基于离群点检测的用电信息采集系统窃电研究

重庆邮电大学

该项目为数据挖掘技术在用电信息采集系统中的应用，利用离群点检测算法处理海量用户用电数据，得到每个用户的窃电嫌疑系数，通过窃电嫌疑系数排序，快速定位窃电嫌疑用户。

- ◆ 通过细致文献调研工作，综合计量原理与窃电实践，选取对窃电现象敏感且便于量化的电气参量作为主要特征，如电压、功率因素、相零线电流不平衡率、电量不平衡率、平均日冻结电量；
- ◆ 利用层次分析法，结合窃电实践与9/9-9/1标度法，给不同特征赋予相应的权重；
- ◆ 采用LOF局部离群因子，混合基于密度离群检测算法与层次分析法获取窃电嫌疑。

相较于传统 LOF 方法，AUC 提升5%；同时受理相关专利1篇。

#### 2015.09 - 2016.02 基于PLC信道相关性的分集合并技术研究

重庆邮电大学

本课题主要为SIMO-PLC系统在低压电力线信道与噪声环境下的分集合并技术误码率性能分析。

- ◆ 通过协方差矩阵Cholesky分解得到具有对应相关特性的瑞利分布变量作为SIMO信道衰落系数；
- ◆ 通过Class-A脉冲噪声概率密度曲线生成随机脉冲噪声作为电力线信道中脉冲噪声；
- ◆ 通过蒙特卡洛仿真的方法分析BPSK调制在EGC等增益合并与MRC最大比合并下的误码率性能。

针对该项目，发表相关中文核心论文1篇，为毕业论文奠定基础。

#### 2014.09 - 2015.08 电网及家用电器对PLC影响研究

重庆邮电大学

该项目为国家电网横向项目，研究并分析不同类型家用电器存在的电磁差模与共模传导骚扰，同时分析其对低压电力线载波通信性能影响。

- ◆ 于现有V型人工电源网络下，研究新型差模传导骚扰测试方法；
- ◆ 研究并测试不同类型家用电器的差模传导骚扰特性，并分析其对PLC通信性能影响。

针对该项目，完成并发表相关SCI论文1篇，见刊于 **IEEE Transaction on EMC**。

## 个人兴趣

2019.7 - 2019.7 阿里云企业云网及跨境代理搭建

南燕保险科技

由于 Salesforce 服务器位于国外，公司 ECS 部署于国内阿里云，两者之间由于网络偶有数据同步失败情况，故通过阿里云企业网，搭建企业云网，通过 Proxy 实现国内 ECS 至 Salesforce 的访问。

- ◆ 负责阿里云企业方案落地，与阿里云对接沟通，确认物理网络方案；
- ◆ 在海外 ECS 通过 tinypoxy 搭建 Http proxy 服务端；
- ◆ 国内 ECS 采用 privoxy 作为 Linux http proxy 客户端，自定义代理规则，从而实现仅 Salesforce 请求通过代理完成，而默认其他请求通过直连国内。

2018.07 - 2018.08 阿里巴巴大数据智能云上编程大赛

在气象业务人员在预报中所广泛参考的多种格点降水预报（集合预报）的结果中，综合考虑各种预报的误差规律，将各个预报的结果融合分析，预测资料覆盖区域的未来24小时降水预报结果。

- ◆ 通过 odps\_sql 完成数据清洗与规整；
- ◆ 通过阿里云端数加平台 MaxCompute、PAI 进行集合降水预报的数据处理、数据计算、机器学习训练与预测。

阿里巴巴大数据智能云上编程大赛（初赛10/1022，复赛15/1022）

2016.12 - 2016.12 网易云计算微专业视频爬虫 ——（Github - wensdayh）

为便于离线式学习，同时将初学的Python用于实践，对网易云计算微专业的公开课视频进行爬取。

- ◆ 通过浏览器调试模式，分析网页跳转机制、Post数据，制定获取视频真实URL流程；
- ◆ 通过 request 库完成对课程相关页面数据请求；
- ◆ 通过 BeautifulSoup 库对获取的 HTML 文件进行解析；
- ◆ 通过 re 库利用正则表达式对HTML文件中字符串过滤、截断等操作；
- ◆ 通过 os 库对视频数据进行本地保存。

完成对视频的爬取，并以有序整理的文档保存于本地磁盘，相关源代码已上传至 Github。

## 荣誉奖励

2014.09-2017.06 校一、二、三等奖学金各1次，优秀硕士学位论文

2010.09-2014.06 校二等奖学金2次，校三等奖学金1次，优秀“三好”学生1次  
“三好”学生1次，校级“知行杯”电子制作大赛三等奖

## 科研成果

[1] Wang Yi, Wen Huian, Hou Xingzhe, et al. Comparison of Differential Mode and Mixed Mode Conducted Emission for Household Appliances in Power Line Communication System [J]. IEEE Transaction on EMC.

[2] 王毅, 温慧安, 张方园, 等. 基于PLC信道相关性的分集合并技术研究[J]. 重庆邮电大学学报（自然科学版）, 2017, 29(1): 9-14.

[3] 基于低压拓扑网络的多节点电力信道响应生成方法, 201610185783.4

[4] 一种基于FPGA的随机脉冲发生器, 201610703140.4

[5] 基于离群点检测的用电信息采集系统实时防窃电检测方法, 201610703224.8

## 自我评价

- 对数据分析、数据挖掘有浓厚兴趣，通过将抽象的算法/模型具体化加深对算法的理解；
- 通信专业本硕 7 年，良好的数学、通信与计算机背景知识；
- 结构化思维，学习能力强，有耐心，能够通过调研和分析，迅速进入新的领域；
- 迎难而上、踏实肯干，对事物保有好奇心，责任心强；
- 具备优秀的发现问题、分析问题与解决问题的能力。