

期望职位：
算法 / 机器学习 / 数据挖掘

孙品

13018986758
sunp@whu.edu.cn
<https://github.com/brightsunp>

工作经历

机器学习工程师	友笑网络 - 杭州	2018 夏 - 至今
---------	-----------	-------------

政要校友信息挖掘

- NLP 项目：前期用 Scrapy 爬取大量网页，中期用 Random-Forest 判别和筛选数据，后期用 NER-CRF 识别实体和抽取信息。
- 在此项目中，我负责技术方案、整体架构的设计，开发了高效、并发的爬虫模块，并且对模型的精度进行了评估和优化。
- 我制定的层级、深度爬取策略，让项目的数据覆盖度高、结构清晰、易维护；并发策略，数倍提升了爬虫效率。

文章推荐系统

- NLP 项目：对于 20w 篇文章，采用无监督学习，用 LDA 构建主题模型，结合 KMeans 进行文章推荐。
- 在此项目中，我负责文本预处理、特征提取、模型训练、参数调优、数据库存储的全流程。
- 我采用 simhash + joblib 实现了文章的增量去重，对于海量数据去重有了较深认识。

软件工程师	华为技术有限公司 - 上海	2017 春 - 2018 夏
-------	---------------	-----------------

5G 持续集成组

- 作为 CI 团队的主要成员，我负责开发和维护代码质量防护工具，采用多进程将运行时间缩短了 60%。
- 在日常的运维中，我开发和重构了常用的工具类，将任务落实为自动化脚本，提升了整个团队的工作效率，荣获团队的“明日之星奖”。

运营商 GUI 项目

- 界面可视化项目：采用 wxPython 框架，便于用户对 license 文件进行操作。
- 在此项目中，我负责数据分析、处理、存储等模块的开发，用 gettext 实现了中英文切换。
- 我参与了团队代码的交叉 review，设计和编写 test-driven 代码，使得项目零缺陷交付。

教育背景

遥感图像分类	中国科学院上海技术物理研究所	2014 秋 - 2017 夏
--------	----------------	-----------------

- 工学硕士，论文课题与图像分类有关。

地理信息系统	武汉大学	2010 秋 - 2014 夏
--------	------	-----------------

- 理学学士，论文课题与城市电子地图有关。

开源项目

-
- [blog & learn](#): 不断充电，记录自己学习和工作的总结。
 - [implement-ML-models](#): 手写 ML 模型，加深对算法原理的理解。
 - [dfld_crawler](#): 一个简单的 Scrapy 爬虫，用来抓取简历判别模型的验证数据。

自我评价

-
- Python 重度用户，熟悉数据结构和算法，以及常用的机器学习模型；
 - 技能标签：Python / NLP / Linux / Node.js / MongoDB 等；
 - 复合专业背景，热爱开源社区，自我驱动式学习，优秀的英文读写能力；
 - 热爱工程师文化，渴望持续的技术积累，看重公司平台和团队氛围。