

WEI JIANG 江伟

|Email : wei.jiang@nyu.edu

|Cell : +86 18802014980

|Personal Web : <https://weijiang2009.github.io/>

|Hackerrank : <https://www.hackerrank.com/weijiang2009>

|Github : <https://www.github.com/weijiang2009>



申请职位 | 数据分析 & 数据挖掘师

自我简述 | 我叫江伟，是土生的 30 岁广州人，此前分别在广州，洛杉矶，深圳和纽约等地从事计算机科学相关的求学，工作和研究。目前聚焦于人工（机器）智能，（大）数据分析与科学。

职业技能 | 深入研究 人工（机器）智能相关学习模型，如回归，决策树，分类，聚类等

深入研究 搜索引擎 性能和质量平衡，及其体系架构

熟悉 数据分析 & 数据科学套件工具，如 Word & Excel, Python & R, Java & Scala etc.

熟悉 大数据平台处理套件工具，如 Hadoop, HBase, Elastic, Spark etc.

熟悉 基本算法 & 数据结构

了解 智能硬件（如航拍无人机，模组式机器人，嵌入式系统）的软硬件开发

流利的 英语，国语和粤语

良好的 沟通协调，解决问题和项目管理 能力

主要研发工作 | **数据分析师 HOWPAY**

2015 年 11 月 – 2016 年 8 月 深圳

主要从事在人工智能背景下，机器学习模型的构建，及相关数据分析&挖掘等研发工作

目标职责 1：使用机器智能（分类 CLASSIFICATION）方法，对亿级文档大数据进行“杂质”识别&去除（如和 文章本身内容关联度不高的广告（由文本，图片组成）即可视为数据石油中的“杂质”）

关键成果：构建（训练）基于有监督学习的机器学习（回归）模型 2 个，并用（约）10 种模型质量评价标准对迭代生成的 模型解析力&预测力 在 4 个数据粒度（整体数据集，公众号，文档和段落）上作量化数据分析，生成数据分析表约 20 份，探索&发现 具良好预测力的数据石油“提纯”配方 1 条，并把此配方应用于整体文档数据集中

目标职责 2：使用机器学习（聚类 CLUSTERING）方法，对中小规模数据集进行类 LDA 主题识别&聚类

关键成果：构建 EM（期望最大化）框架 1 个；并基于此框架&无监督机器学习，训练混合高斯模型 1 个和 LDA 主题模型 1 个；由数据和模型的交互，以触发机器学习模型迭代&改善（约）15 次，并用（约）5 种模型质量评价标准分别对模型解析力&预测力在 2 数据粒度（整体数据集，单条数据样本）上作量化数据分析，生成数据分析表约 5 份，为组织今后关于无监督学习的后续研发工作提供理论和实验原型的支持

目标职责 3：使用机器学习（自然语言处理 NLP）方法，对中小规模数据集进行中文分词预研

关键成果：设计&实现 1 个关于中文分词的数据可视化工具（主要包含 1 高效率（正常人工审核速度约为 1000 条/小时（注：此处假定数据样本为社交类语言短文本集，平均长度约为 100 中文字））的机器学习&人机审核交互界面），为组织关于中文自然语言处理及配套下游相关业务（如搜索&推荐）提供理论和实验原型的支持

软&硬件工程师 DJI

2015 年 8 月 – 2015 年 10 月 深圳

目标职责&关键成果：DJI 无人机（机器学习）平台 SDK 文档的整体策划和撰写，包括入门，开放协议文档等，具体请参阅 [HTTPS://GITHUB.COM/DJI-SDK/](https://github.com/DJI-SDK/) 中的中英文文档。

助理研究与教学员 纽约大学工程学院 网络搜索与信息挖掘技术实验室

2011 年 9 月 – 2015 年 3 月 纽约 & 上海

主要从事在信息检索背景下，搜索引擎性能和质量调优，机器学习和数据挖掘的研究工作

目标职责 1：研究各种搜索系统静态倒排索引的剪枝（Static Inverted Index Pruning）算法

关键成果：在保持搜索质量无明显下降的前提下，使倒排索引的大小缩减为原索引大小的 5%（详见论文 1）

目标职责 2：研究各种倒排索引分层（Inverted Index Tiering）方法

关键成果：在保持搜索质量无明显降低的前提下，使实验系统整体吞吐量（Query Throughput）提高 80%，用户请求平均延迟（Query Latency）减少 10%

目标职责 3：研究用户三维检索（关键字，时间和空间）在搜索系统中的处理算法

关键成果：对于 60% 的用户检索请求，在保持搜索质量无明显降低的前提下，有 3-5 倍的处理速度提升（详见论文 2）

目标职责 4&关键成果：助理教学若干计算机科学课程，如操作系统

研究程序员 南加州大学工程学院 机器人和嵌入式系统中心 多形态机器人实验室

2009 年 9 月 – 2011 年 5 月 洛杉矶，加利福尼亚州，美国

目标职责：自适应，自重构机器人模组的长距离传感和规划，及自主对接(Autonomous Docking)的算法设计和实现

关键成果：在大于 60%的实验场景下，两个机器人模组均能通过 IR 通讯并自主规划前期对接行为，相互靠近并为最终对接作准备（详见研发视频部分 1 和 2）

项目负责人 广东工业大学 计算机学院 嵌入式系统与计算智能实验中心

2007 年 9 月 – 2009 年 5 月 广州市，广东省，中国

目标职责：研发一个手持型室内墙埋电线检测仪，能无伤探测场内电线走向，主要负责其嵌入式系统软件信号处理模块和物理传感器模块的研发工作

关键成果：详见产品视频介绍 3 和相关专利授权 1，2，3，4

教育背景

南加州大学 工程学院 计算机科学系

硕士 计算机科学 2009 年 9 月 – 2011 年 5 月，洛杉矶，加利福尼亚州，美国

绩点：3.7/4.0

主要课程：机器学习，数据挖掘，自然语言处理，分布式系统，机器人

研究兴趣：多形态机器人系统自适应和自主算法



广东工业大学 计算机学院

学士 计算机科学 2005 年 9 月 – 2009 年 5 月，广州市，广东省，中国

绩点：88/100 排名：4/800

主要课程：算法，数据结构，操作系统，嵌入式系统，数据库，统计学

研究兴趣：嵌入式系统



发明和实用新型专利

1. 一种非电流互感器式工频弱电电磁信号传感器 CN 201425627 Y

发明人：江立锋，江纬，何元烈，余永权 [Google 专利链接](#)

2. 一种室内墙埋电线无伤检测仪 CN 101539637 A

发明人：江立锋，江纬，何元烈，余永权 [Google 专利链接](#)

3. 非电流互感器式工频弱电电磁信号传感器 CN 101539636 A

发明人：江立锋，江纬，何元烈，余永权 [Google 专利链接](#)

4. 室内墙埋电线无伤检测仪 CN 201425626 Y

发明人：江立锋，江纬，何元烈，余永权 [Google 专利链接](#)

主要论文

1. **Improved Methods for Static Index Pruning** 和 J.Rodriguez, T.Suel. (In Proceedings of IEEE BigData 2016) [链接 1](#)
2. **What, Where, and When: Geotemporal Text Search in Twitter** 和 S.Nepomnyachiy, B.Gelley, T.Minkus. In Proceedings of Geographic Information Retrieval 2014 [链接 1](#)

主要奖项和荣誉

1. 第十一届挑战杯全国大学生课外学术科技作品竞赛 国家赛 三等奖
2. 第十届挑战杯全国大学生课外学术科技作品竞赛 广东省赛 特等奖
3. 2008 年中国国家奖学金
4. 南加州大学第一届机器人大赛第二名

相关研究视频

1. SuperBot 机器人短距离趋近对接视频：[Youtube 链接 1](#) [实验室链接 2](#)
2. SuperBot 机器人长距离传感和规划视频：[Youtube 链接 1](#) [实验室链接 2](#)
3. 室内墙埋电线无伤检测仪介绍视频：[Youtube 链接](#)

推荐/联系人

(根据需要提供)