吴嘉澍

男 1997.06.13 邮箱 [jiashu@student.unimelb.edu.au](mailto:jiashu@student.unimelb.edu.au) 微信 jiashu613 [linkedin.com/in/jiashu-wu](https://www.linkedin.com/in/jiashu-wu)

籍贯 北京 海淀 电话 17801323125 Skype live:davidjiashuwu [github.com/jiashuwu](https://github.com/jiashuwu)

# 教育背景

**墨尔本大学** 信息技术硕士人工智能2019.02 – 2020.12

均分88.1 专业前2% 全科First Class Honours 入选2019 Dean’s Honours List

**悉尼大学** 理科学士计算机科学 金融数学与统计学 双专业2016.02 – 2018.12

均分86.5 专业前2% 计算机科学均分85 金融数学与统计学均分88 入选2017 & 2018 Dean’s Honours List

**北京理工大学** 软件工程2015.08 – 2016.01

软件工程专业 2016年转学至悉尼大学就读

# 获奖经历

2019 Dean’s Honours List 墨尔本大学工程系

2018 Dean’s List of Excellence in Academic Performance 悉尼大学理学系

2017 Dean’s List of Excellence in Academic Performance 悉尼大学理学系

# 研究经历

2020 **中国科学院深圳先进技术研究院** 研究实习生 2020.09 – 至今

中国科学院深圳先进技术研究院数字所云计算技术研究中心

导师：王洋教授

* 探究云计算技术中的一些重点问题，如动态资源分配、资源分区、基于机器学习的并行文件系统性能优化、存储系统性能优化等课题。围绕这些课题进行探究，现已完成论文4篇，撰写发明专利8项。
* 参与多想国家级及省级云计算研究项目。

2019 **北京理工大学计算机学院** 研究实习生2019.11 – 2020.05

北京理工大学计算机科学与技术学院计算机视觉实验室

课题：Simultaneous Semantic Alignment Network for Heterogeneous Domain Adaptation

导师：李爽副教授

* 探究挑战性大的HDA问题，难点在于特征表达与模态的异构性。通过不断试验分析，成功克服HDA模型标签准确率低等问题。
* 探索不同的网络结构，试验不同的迁移方法。通过主动思考，最终有效地使用知识蒸馏、中心点匹配、标签精细化等方法，实现对于模型迁移能力的优化增强。成功地使模型在多个HDA数据集上比现有HDA方法均有约1~6%的预测准确率提升。
* 使用PyTorch深度学习框架实现模型，并利用Git维护管理不同版本的模型代码。代码可读性高，注释清晰，整洁易懂。
* 在有限时间内高质量完成9千字英文论文。论文及模型创新性获审稿人一致认可，并已被CCF-A类会议ACM MM’2020接收。

2019 **墨尔本大学计算与信息系统学院** 研究开发工程师 2019.07 – 2019.11

墨尔本大学计算与信息系统学院

课题：Learning to rank with small set of ground truth data

导师：Professor Rui Zhang

* 负责开发学者搜索平台。不惧难点，在规定时间内高效试验矩阵分解及知识库两种解决方案，成功解决排序数据不足的难点。
* 运用Python (NLTK，spaCy) 对40万条论文数据及词条数约为2千万的Wiki和MAG词库进行数据预处理，提取与整合。
* 缜密分析并对比多种算法的优劣，包括LSA，NMF及深度推荐系统算法NFM。高效处理、挖掘大量数据，利用数据量约20亿的MAG Knowledge Base辅助实现搜索结果的推荐及排序，成功解决了搜索词未在论文中出现等难点。
* 高质量撰写12000字近50页的报告，详细展示了数据获取，数据处理，模型试验结果，最终模型方案等内容，并做汇报。

2018 **悉尼大学信息技术学院** Dalyell Scholar项目 2018.03 – 2018.07

悉尼大学信息技术学院

课题：Artificial Intelligence for medical screening using Graphonomics

导师：Associate Professor Simon Poon

* 阅读文献，探究通过画图笔迹数据判断帕金森病的方法
* 在悉尼大学MHTI实验室负责编写收集用户绘制图形的笔迹坐标及触点压力数据的安卓App，数据采用SQLite3进行高效存储。
* 应用可靠易用，帮助实验室顺利完成数据采集并用于帕金森诊断的研究，后续研究进展顺利。

# 论文发表

2020 Jiashu Wu, Hongbo Wang, Hao Dai, Chengzhong Xu and Yang Wang  , “Research on machine learning-based performance optimization of dynamic partitioned parallel file system”, Journal of Integration Technology, 2020, 9(6): pp 71-83. doi: 10.12146/j.issn.2095-3135.20200901001

2020 Shuang Li , Binhui Xie, Jiashu Wu, Ying Zhao, Chi Harold Liu and Zhengming Ding, “Simultaneous Semantic Alignment Network for Heterogeneous Domain Adaptation”, ACM International Conference on Multimedia (ACM MM), 2020. (Accepted, **CCF A Conference**, link: [arxiv.org/abs/2008.01677](https://arxiv.org/abs/2008.01677), [dl.acm.org/doi/10.1145/3394171.3413995](https://dl.acm.org/doi/10.1145/3394171.3413995))

2020 Jiashu Wu, Yang Wang  , Jinpeng Wang, Hekang Wang, Taorui Lin and Chengzhong Xu, “How does SSD cluster perform for distributed file systems: An empirical study”. (Under review at The Computer Journal)

2020 Huaxiao Rao, Jiashu Wu, Xiaopeng Fan and Yang Wang  , “Multidimensional application recommender system based on user feature hierarchical clustering with user behaviour information”. (Under review at the Journal of Integration Technology)

# 发明专利

2021 一种基于二次预过滤与联合分区的theta-join优化算法 （中国专利已提交）

2021 一种利用用户聚类特征与用户情境信息的基于概率的应用推荐系统 （中国专利已提交）

2020 在线场景的资源分配方法、装置及电子设备 （CN202011428352.9）

2020 一种动态资源分区方法 （CN202011384022.4）

2020 搜索模型训练方法、装置、终端设备及存储介质 （CN202011403845.7）

2020 文献搜索方法、装置及电子设备 （CN202011433146.7）

2020 lock-free的分布式死锁避免方法及装置、计算机设备及可读存储介质 （CN202011438337.2）

2020 一种数据流连接优化方法、系统、终端以及存储介质 （CN202011435327.3）

# 实习经历

2020 **墨尔本大学 Melbourne eResearch Group 软件工程师** 2020.03 – 2020.06

* 开发一款用于会议对话识别分析的安卓App，此应用将被墨尔本大学图书馆用于进行会议参与度的研究。
* 应用使用了Material Design设计组件，以及Google机器学习语音识别API。
* 为项目提供了高效整洁的代码以及可读性高的文档。

2017 **悉尼大学信息技术系助教** 2017.03 – 2017.11

* 担任悉尼大学信息技术系INFO1003网页设计、INFO1103Java程序设计、INFO1105数据结构课的助教工作。
* 协助老师为上百名同学答疑，展现良好的沟通交流能力与英语水平，连续两学期获得学院优秀助教荣誉。

# 项目经历

**Kaggle推特文本作者归属分类** 2019.07 – 2019.09

* 利用Python对30万条推特进行数据清洗和预处理。开展特征工程，并进行特征选择。利用SMOTE算法解决样本不均衡的问题。
* 负责使用SVM，随机森林，TextCNN，TextRNN，FastText等分类算法，成功将准确率提升至30%，排名20 (约200支队伍)。

**Bitbox 分布式文件系统** 2019.03 – 2019.06

* 使用Java编写一个可以在客服端见同步文件的非中心式文件系统。
* 文件系统客户端之间可以通过共有-私有密钥实现对通讯内容的加密。

# 专业技能

* 熟练使用 Python (NumPy, NLTK, Matplotlib, Sklearn, PyTorch, Pandas, PyMySQL等) ，Java (JDBC)，SQL，R(假设检测，统计分析，作图，时间序列分析等)，可使用C，Haskell。自学Hadoop，Spark及Hive基础以及Power BI可视化工具。
* 熟练使用 Windows操作系统，Microsoft Excel, PPT, Word, Visio, 等软件，熟悉Linux操作系统。可使用Git 进行版本管理。
* 熟悉数据清洗，数据处理，数据建模等数据处理分析流程。熟悉聚类，神经网络等多种机器学习、深度学习算法。
* 主要课程：机器学习，数据分析，统计学习，数据库，数据挖掘，信息检索，自然语言处理，统计建模，统计测试等。

# 语言能力

* 雅思学术类 (总分7.0，阅读 8.0，听力 8.0，口语 6.0，写作6.0)。
* 大学英语四级 (总分665，阅读满分)。
* 在澳大利亚五年学习生活经历，熟练适应英语交流环境，能够熟练使用英语进行汇报及写作，具有良好的沟通交流表达能力。

# 其他经历

* 具有极强的学习能力。从零自学安卓开发，业余时间编写的倒数日记事本App已上架华为、阿里等应用市场，已有几千次下载量。
* 担任北京理工大学学生会体育部部员，责任心强，参与筹备上千人参与的校运会及长跑。协调沟通场地等多部门，活动顺利举办。