David Jiashu Wu

Male | DoB: 13th Jun 1997
Email: jiashu@student.unimelb.edu.au
LinkedIn: linkedin.com/in/jiashu-wu/
GitHub: github.com/jiashuwu

EDUCATION BACKGROUND

The University of Melbourne 2019.02 – 2020.07

Master of Information Technology (with Distinction), with WAM **88.1 First Class Honours**

Stream: **Artificial Intelligence**

Major curriculums: Distributed Systems, Database System, Machine Learning, Natural Language Processing, Artificial Intelligence, Information Retrieval, Recommender Systems etc.

The University of Sydney 2016.02 – 2018.12

Bachelor of Science, weighted average mark **86.5 High Distinction**

Top 2% in Faculty of Science, entered **Talented Student Program**

Major: **Computer Science**, weighted average **85 High Distinction**

Major curriculums: Operating Systems, Database, Algorithm and Complexity, Computer network, Data Analytics, Machine Learning, Human-Computer Interaction, Website Design, Project Management etc.

Major: **Financial Mathematics and Statistics**, weighted average **88 High Distinction**

Major curriculums: Statistical Models, Data Analysis, Financial Mathematics, Computational Science, Statistical Tests, Stochastic Processes, Time Series Analysis, Optimization, Statistical Learning etc.

SCHOLARSHIP

2019 **Dean's Honours List** School of Engineering, University of Melbourne

2018 **Dean's List of Excellence in Academic Performance** Faculty of Science, University of Sydney

2017 **Dean's List of Excellence in Academic Performance** Faculty of Science, University of Sydney

RESEARCH EXPERIENCE

2020 **Research Intern**

Centre for Cloud Computing, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

Supervisor: Professor Yang Wang, Centre for Cloud Computing, SIAT, Chinese Academy of Sciences

2020 **Research Intern**

School of Computer Science and Technology, Beijing Institute of Technology

Project Title: Simultaneous Semantic Alignment Network for Heterogeneous Domain Adaptation

Supervisor: Associate Professor Shuang Li, School of Computer Science and Technology

2019 **Computing Research Project**

School of Computing and Information Systems, University of Melbourne

Project Title: Learning to rank with small set of ground truth data

Supervisor: Professor Rui Zhang, School of Computing and Information Systems

2018 **Talented Student Program** (a.k.a. Dalyell Scholar Program)

School of Information Technology, University of Sydney

Project Title: Artificial Intelligence for medical screening using Graphonomics (App Development and Testing)

Supervisor: Associate Professor Simon Poon, School of Information Technology

PUBLICATIONS

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, **CORE Conference Rank A***, **CCF A Conference**, link: arxiv.org/abs/2008.01677, dl.acm.org/doi/10.1145/3394171.3413995)

PATENTS

2020 An machine learning based multi-scenario dynamic online resource allocation algorithm (China Patent under review)

INTERNSHIP EXPERIENCE

2020 **Software Engineer at Melbourne eResearch Group** 2020.03 – 2020.06

- Develop a meeting speaker diarization Android App. The app will then be used by UniMelb Library for research purposes.
- The app utilizes Material Design Components, as well as Google ML Speech API. Well-commented code and the documentation is publicly available.

2019 **Research Intern at Beijing Institute of Technology** 2019.11 – 2020.03

- Tackle Heterogeneous Domain Adaptation (HDA) problem, where the source and the target domains have heterogeneous feature representations, and may also come from diverse modalities.
- Utilize knowledge distillation to transfer the semantic knowledge between two domains. Together with explicit semantic alignment, it enhances the adaptability of the purposed model.
- Leverage the three-prototype alignment to explicitly transfer the semantic knowledge across domains. To mitigate the transferability degradation caused by false pseudo-labels, geometric similarity is used to refine the pseudo-label assignment. The model yields the state-of-the-art performance on several HDA datasets (NUSTAG – ImageNet etc.) and outperforms other HDA works by 1~6%.
- Complete the paper in high-quality and the paper has been accepted by ACM MM'2020. The code is written using PyTorch framework.

2019 **Research and Development Engineer at Profware** 2019.7 – 2019.11

- Develop an academia searching platform. Challenges including limited amount of ground truth ranking results, and the searching platform should be able to search for researchers even if the query keywords don't explicitly appear in the researcher's papers.
- Utilize Python NLTK and spaCy to pre-process the publication datasets with approximately 400k research papers and publications, and a term dataset retrieved from Wiki and MAG with 20 million entries. The pre-processing steps including sentence tokenization, lemmatization, etc. BoW model is then used to build matrices, and Learning-to-Rank techniques like Pseudo Relevance Feedback is leveraged to transform the matrices.
- Several algorithms are experimented and analyzed, including LSA, Non-negative Matrix Factorization and deep recommender system algorithm Neural Factorization Machine. The MAG Knowledge Base with 20 billion entries is used to assist the result recommendation and ranking.

2017 **Mentor at School of Information Technology, University of Sydney** 2017.3 – 2017.11

- Mentor for course INFO1003 Website Design, INFO1103 Java Programming and INFO1105 Data Structures.
- Help tutor to answer questions, share my experience with students, demonstrate an excellent communication skill and interpersonal skill.

PROJECTS

Kaggle Twitter Author Attribution 2019.07 – 2019.09

- Use Python to clean and processes 300k Tweets, conduct feature engineering and feature selection. Utilize SMOTE algorithm to solve the sample imbalance problem.
- Experiment using algorithms including SVM, RF, TextCNN, TextRNN and FastText. Successfully achieve 30% classification accuracy, and rank 20/200 on Kaggle.

BitBox Distributed File System 2019.03-2019.06

- Use Java to implement a distributed file system, capable of synchronizing file directories between peers in a decentralized network.
- Clients can securely communicate with the BitBox peers, using public-private key cryptography.

TECHNICAL ABILITY

Programming - Mainly use Python and Java, able to use MySQL, C, R and MATLAB.

Technical Skills – Database, Data Mining, Data Analysis, Statistics, Machine Learning, Distributed Systems.

LANGUAGE ABILITY

Chinese	Native Speaker
English	IELTS Academic 7.0, with reading 8.0, listening 8.0, speaking 6.0 and writing 6.0
	Lived and studied in Sydney Australia for three years and Melbourne Australia for two years.