

David Jiashu Wu

PERSONAL INFORMATION

Male | DoB: 13th Jun 1997 | Beijing Haidian
LinkedIn: [linkedin.com/in/jiashu-wu/](https://www.linkedin.com/in/jiashu-wu/)

Email: js.wu@siat.ac.cn | jiashu@student.unimelb.edu.au
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**, top 2%

Stream: **Artificial Intelligence**

Major curriculums: Distributed Systems, Database System, Machine Learning, Natural Language

Processing, Artificial Intelligence, Deep Learning, 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.

Beijing Institute of Technology 2015.08 – 2016.01

Major in software engineering, transferred to University of Sydney in 2016

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 **Student Intern at Chinese Academy of Sciences** 2020.09 - Current

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

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

- Work on challenging cloud computing topics including dynamic resource allocation, resource partitioning, machine learning-based parallel file system performance optimization, etc.
- Conduct thorough research, submitted 4 papers and 8 patents.
- Participated in several Chinese national and provincial cloud computing projects.

2020 **Student Intern at Beijing Institute of Technology** 2019.11 – 2020.03

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

- 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 University of Melbourne** 2019.07 – 2019.11

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

- 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.

2018 **Talented Student Program** (a.k.a. Dallyell Scholar Program) 2018.03 – 2018.07

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

- Carry out research on popular methods that utilize drawing to detect Parkinson Disease
- Develop an Android App that is capable to collect user's drawing trace data on various different pattern templates.

PUBLICATION

- | | |
|------|--|
| 2020 | Yi Shu 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 , 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) |

PATENT

- | | |
|------|---|
| 2021 | A theta-join optimization algorithm based on double pre-filtering and associated partitioning mechanism (China Patent Submitted) |
| 2021 | A probabilistic application recommender system based on user feature clustering and user behaviour information (China Patent Submitted) |
| 2020 | A machine learning based multi-scenario dynamic online resource allocation algorithm (CN202011428352.9) |
| 2020 | A deadlock-free high concurrency dynamic resource partitioning algorithm (CN202011384022.4) |
| 2020 | A publication retrieval model's training method, mechanism, end device and storage medium (CN202011403845.7) |
| 2020 | A knowledge tree-based publication retrieval algorithm, mechanism and end device (CN202011433146.7) |
| 2020 | A lock-free distributed deadlock avoidance algorithm, and its related mechanism, computer device and readable storage medium (CN202011438337.2) |
| 2020 | An online data stream theta-join optimization algorithm, system, end device and storage medium (CN202011435327.3) |

INTERNSHIP EXPERIENCE

- | | | |
|---|---|-------------------|
| 2020 | Software Engineer at Melbourne eResearch Group | 2020.03 – 2020.06 |
| <ul style="list-style-type: none">• 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. | | |
| 2017 | Mentor at School of Information Technology, University of Sydney | 2017.3 – 2017.11 |
| <ul style="list-style-type: none">• 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. | | |

PROJECT

- | | |
|---|-------------------|
| Kaggle Twitter Author Attribution | 2019.07 – 2019.09 |
| <ul style="list-style-type: none">• 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 |
| <ul style="list-style-type: none">• 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;
CET4 overall 665, achieved full mark in reading;
Lived and studied in Sydney Australia for three years and Melbourne Australia for two years. |