David Jiashu Wu

PERSONAL INFORMATION

Male | DoB: 13th Jun 1997 | Beijing Haidian
Email: js.wu@siat.ac.cn | jiashu@student.unimelb.edu.au
LinkedIn: linkedin.com/in/jiashu-wu/
GitHub: github.com/jiashuwu
WeChat: jiashu613

EDUCATION BACKGROUND

The University of Melbourne 2019.02 – 2020.07

Master of Information Technology (with Distinction), weighted average 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, Optimisation, Statistical Learning etc.

Beijing Institute of Technology 2015.08 – 2016.01

Major in Software Engineering, transferred to the 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 optimisation, storage system performance evaluation etc.
- Conduct thorough researches, completed 4 papers and 12 patents (including 4 PCT patents).
- Participate in several Chinese national and provincial cloud computing projects.

2019 **Student Intern at Beijing Institute of Technology** 2019.11 – 2020.05

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.
- Utilise 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, the geometric similarity is used to refine the pseudo-label assignment. The model yields 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 as poster and published in ACM MM'2020 (CCF A). 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 a limited amount of ground truth ranking data, 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.
- Utilise 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 tokenisation, lemmatisation, etc. The 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 analysed, including LSA, Non-negative Matrix Factorisation and deep recommender system algorithm Neural Factorisation Machine. The MAG Knowledge Base with 20 billion entries is used to assist the result recommendation and ranking.

2018 **Dalyell Scholar Program at University of Sydney** (Talented Student 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 utilise drawing patterns to detect the Parkinson Disease.
- Develop an Android App that is capable to collect user's drawing trace data on several different pattern templates.
- Manage data using SQLite3, the data collected helps the lab to carry out further research on Parkinson Disease diagnosis.

PUBLICATION

2020 Jiashu Wu, Hongbo Wang, Hao Dai, Chengzhong Xu and Yang Wang ✉, "Research on Machine Learning-based Performance Optimisation 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. (**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, **CCF B**)

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)

2020 Jiashu Wu, Yang Wang ✉, Ziyue Hu, Xiaopeng Fan, Kejiang Ye and Chengzhong Xu, "Towards Faster Theta-join: A Pre-filtering and United Partitioning Approach". (Under review at the IEEE Transactions on Big Data, **CCF C**)

PATENT

2021 A Theta-join Optimisation Algorithm based on Double Pre-filtering and United 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, PCT/CN2020/139560)

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, PCT/CN2020/140016)

2020 A Knowledge Tree-based Publication Retrieval Algorithm, Mechanism and End Device (CN202011433146.7, PCT/CN2020/139264)

2020 A Lock-free Distributed Deadlock Avoidance Algorithm, and its Related Mechanism, Computer Device and Readable Storage Medium (CN202011438337.2, PCT/CN2020/139345)

2020 An Online Data Stream Theta-join Optimisation Algorithm, System, End Device and Storage Medium (CN202011435327.3)

INTERNSHIP EXPERIENCE

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

- Develop a meeting speaker diarisation Android App. The app will then be used by the UniMelb Library for research purposes.
- The app utilises 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** 2017.03 – 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.

PROJECT

Kaggle Twitter Author Attribution 2019.07 – 2019.09

- Use Python to clean and process 300k Tweets, conduct feature engineering and feature selection. Utilise SMOTE algorithm to solve the sample imbalance problem.
- Experimenting 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 synchronising file directories between peers in a decentralised network.
- Clients can securely communicate with BitBox peers, using public-private key cryptography.

TECHNICAL ABILITY

Programming – Mainly use Python (NumPy, NLTK, Matplotlib, Sklearn, PyTorch, Pandas etc.), Java, SQL, R (hypothesis testing, statistical analysis, plotting, time series analysis etc.), able to use C, Haskell, and MATLAB.

Technical Skills – Database, Data Mining, Data Analysis, Statistics, Machine Learning, Deep 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 a full mark in reading;

Lived and studied in Sydney Australia for three years and Melbourne Australia for two years.