

Curriculum Vitae

Lan Du

Lecturer in Data Science
Deputy Course Director of Data Science
Faculty of Information Technology
Building 63 - 25 Exhibition Walk
Monash University Clayton Campus
Wellington Road, Clayton
VIC 3800

Office phone: (61) 3 9905 9971
Mobile: (61) 433018598
Email: lan.du@monash.edu

Research Interests

His research interests focus on statistical modelling and learning, which broadly cover statistical machine learning, natural language processing, data mining, and social network analysis. He is particularly interested in the following research areas: topic modeling, nonparametric Bayesian methods, probabilistic matrix/tensor factorisation, deep Bayesian models, statistical ranking models, etc.

Education

1. B.S. Information Technology,

University: **the Flinders University of South Australia**

When: Jul. 2004 – Jul. 2006.

Degree awarded: Jul, 2006

2. B.S. Information Technology with **1st class honours**

University: **The Australian National University**

When: Feb. 2007 – Dec. 2007

Thesis title: *Ontology-Driven Information Retrieval for Digital Forensics*

Honours project supervisor: Dr. [Huidong Jin](#)

Degree awarded: December, 2007

3. Ph.D. in Computer Science

University: **The Australian National University**

When: Aug. 2008 - Dec. 2011

Thesis title: *Nonparametric Bayesian Methods for Structured Topic Models*

Supervisors: Prof. [Wray Buntine](#) and Dr. Huidong Jin

Degree awarded: December, 2012

Employment History

Lecturer in Data Science (Faculty member),

Employer: Faculty of Information Technology, **Monash University**

When: Sep. 2015 - Now

Post-Doc Research fellow

Employer: Department of Computing, **Macquarie University**

When: Dec. 2011 - Aug. 2015

Supervisor: Prof. [Mark Johnson](#)

Awards

Monash Engineering-IT-Science ECR Interdisciplinary Research (IDR) Seed fund 2017 (AUD 9,000)

Project: "Explore life stories of academics with sequential models"

Principle investigator: Dr. **Lan Du**

[Google Natural Language Understanding focused awards 2013](#), (USD225,000).

Project: "Generative Models of Life Stories"

Chief investigators: Professor Mark Johnson (Macquarie University, Australia), Dr. **Lan Du**, and Professor Wray Buntine (Monash University, Australia)

Scholarships

ANU-NICTA PhD scholarship, 2008–2011.

NICTA PhD Supplementary scholarship, 2008–2011.

NICTA PhD Assignment scholarship, 2008–2011.

ANU summer research scholarship, 2008.

ANU-NICTA honours scholarship, 2007.

Publications

Book Chapters

1. Zahraa S Abdallah, **Lan Du**, Geoffrey I Webb, "Data Preparation", in **Encyclopedia of Machine Learning and Data Mining**, pages 1-11, 2016

Journal Articles

10. He Zhao, **Lan Du**, Wray Buntine, Gang Liu, "Leveraging External Information In Topic Modelling", *Knowledge and Information Systems* 2018, to appear.
9. Jianping Gou, Yong Xu, David Zhang, Qirong Mao, **Lan Du**, Yongzhao Zhan, "Two-Phase Linear Reconstruction Measure-Based Classification for Face Recognition", *Information Sciences*, Volumes 433-434, Pages 17-36

8. Kar Wai Lim, Wray Buntine, Changyou Chen and **Lan Du**, "Nonparametric Bayesian Topic Modelling with the Hierarchical Pitman-Yor Processes", *International Journal of Approximate Reasoning*, Volume 78, page 172-191, 2016 (Impact factor 2.845)
7. Dat Quoc Nguyen, Richard Billingsley, **Lan Du**, and Mark Johnson, "Improving Topic Models with Latent Feature Word Representations", *Transactions of the Association for Computational Linguistics*, Volume 3, page 299-313, 2015
6. Changyou Chen, Wray Buntine, Nan Ding, Lexing Xie, and **Lan Du**, "Differential Topic Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 37, Issue 2, page 230-242, 2015. (Impact factor 8.329)
5. **Lan Du**, Wray Buntine, Huidong Jin, Changyou Chen, "Sequential Latent Dirichlet Allocation", *Knowledge and Information Systems*, Volume 31, Number 3, 475-503, 2012 (Impact factor 2.004)
4. Jianping Gou, Zhang Yi, **Lan Du**, Taisong Xiong, "A Local Mean-Based K-Nearest Centroid Neighbor Classifier", *The Computer Journal*, Volume 55, Issue 9, 1058-1071, 2012 (Impact factor 0.711)
3. Jianping Gou, **Lan Du**, Taisong Xiong, "Weighted K-Nearest Centroid Neighbor Classification", *Journal of Computational Information Systems* 8: 2 (2012) 851-860,
2. Jianping Gou, **Lan Du**, Yuhong Zhang, Taisong Xiong, "A new distance-weighted k-nearest neighbor classifier", *Journal of Information and Computational Science*, 9: 6 (2012) 1429-1436,
1. **Lan Du**, Wray Buntine, Huidong Jin, "A Segmented Topic Model based on the Two-parameter Poisson-Dirichlet Process", *Machine Learning Journal*, Volume 81, Number 1, page 5-19, 2010 (Impact factor 1.848)

Conference papers

23. He zhao, **Lan Du**, Wray Buntine, Minyuan Zhou "Inter and Intra Topic Structure Learning with Word Embeddings", to appear in **ICML 2018**
22. Yuan Jin, **Lan Du**, Ye Zhu, Mark Carman "Leveraging Label Category Relationships in Multi-class Crowdsourcing", to appear in **PAKDD 2018**
21. He Zhao, Piyush Rai, **Lan Du**, Wray Buntine "Bayesian Multi-label Learning with Sparse Features and Labels, and Label Co-occurrences", to appear in **AISTATS 2018**.
20. He Zhao, **Lan Du**, and Wray Buntine "A Word Embeddings Informed Focused Topic Model", **ACML 2017**, pages 423-438
19. He Zhao, **Lan Du**, Wray Buntine, and Gang Liu "MetaLDA: a Topic Model that Efficiently Incorporates Meta information", **ICDM 2017**, pages 635-644
18. He Zhao, **Lan Du** and Wray Buntine, "Leveraging Node Attributes for Incomplete Relational Data", Proceedings of the 34th International Conference on Machine Learning (**ICML 2017**), pages 4072-4081, 2017
17. Shervin Malmasi, Mark Dras, Mark Johnson, **Lan Du** and Magdalena Wolska "Unsupervised Text Segmentation Based on Native Language Characteristics", Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (**ACL Volume 1: Long Papers**), pages 457-1469, 2017
16. Huiwen Liu, Jiajie Xu, Kai Zheng, Chengfei Liu, **Lan Du**, Xian Wu, "Semantic-aware Query Processing for Activity Trajectories", Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (**WSDM**), Pages 283-292, 2017

15. Jianping Gou, **Lan Du**, Keyang Cheng, Yingfeng Cai, "Discriminative sparsity preserving graph embedding", IEEE Congress on Evolutionary Computation (CEC), Pages 4250-4257, 2016
14. **Lan Du**, Anish Kumar, Mark Johnson, and Massimiliano Ciaramita, "Using Entity Information from a Knowledge Base to Improve Relation Extraction", Australasian Language Technology Association Workshop (ALTA) 2015, pages 31-38
13. Zhendong Zhao, **Lan Du**, Benjamin Börschinger, John K Pate, Massimiliano Ciaramita, Mark Steedman and Mark Johnson, "A Computationally Efficient Algorithm for Learning Topical Collocation Models", page 1460-1469, **ACL2015** (corresponding author)
12. **Lan Du**, John K. Pate and Mark Johnson, "Topic Segmentation with An Ordering-Based Topic Model", page 2232-2238, **AAAI 2015**
11. **Lan Du**, John K Pate, and Mark Johnson "Topic Models with Topic Ordering Regularities for Topic Segmentation", page 803-808, **ICDM 2014**
10. Youliang Zhong, **Lan Du**, Jian Yang "Learning Social Relationship Strength via Matrix Co-Factorization with Multiple Kernels", The 14th International Conference on Web Information System Engineering (**WISE**), page 15-28, 2013
9. **Lan Du**, Wray Buntine, and Mark Johnson, "Topic segmentation with a structured topic model," **NAACL**, pages 190-200, 2013
8. Huidong Jin, Lijiu Zhang, **Lan Du**, "Semantic Title Evaluation and Recommendation Based on Topic Models", The 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD**), pages 402-413, 2013
7. **Lan Du**, Wray Buntine, Huidong Jin, "Modelling Sequential Text with an Adaptive Topic Model", Empirical Methods in Natural Language Processing (**EMNLP**), page 535-545, 2012
6. Changyou Chen, **Lan Du**, Wray Buntine, "Sampling Table Configurations for the hierarchical Poisson-Dirichlet Process", The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (**ECML-PKDD**), page 296-311, 2011
5. **Lan Du**, Wray Buntine, Huidong Jin, "Sequential Latent Dirichlet Allocation: Discover Underlying Topic Structures within a Document", Proceedings of the 2010 IEEE International Conference on Data Mining (**ICDM**, in top 10 papers), page 148-157, 2010.
4. Wray Buntine, **Lan Du**, Petteri Nurmi, "Bayesian Networks on Dirichlet Distributed Vectors", Proceedings of the Fifth European Workshop on Probabilistic Graphical Models (**PGM-2010**), page 33-40, 2010.
3. **Lan Du**, Huidong Jin, Olivier Y. de Vel, Nianjun Liu, "A Latent Semantic Indexing and WordNet based Information Retrieval Model for Digital Forensics", Proceedings of IEEE International Conference on Intelligence and Security Informatics (**ISI**), page 70-75, 2008.
2. Phan Thien Son, **Lan Du**, Huidong Jin, Olivier Y. de Vel, Nianjun Liu, Terry Caelli, "A Simple WordNet-Ontology Based Email Retrieval System for Digital Forensics", Pacific Asia Workshop on Cybercrime and Computer Forensics (ISI workshop), page 217-228, 2008.
1. Junlei Song, Dianhong Wang, Nianjun Liu, Li Cheng, **Lan Du**, Ke Zhang, "Soil Moisture Prediction with Feature Selection Using a Neural Network", Proceedings of International Conference on Digital Image Computing: Techniques and Applications (**DICTA**), page 130-136, 2008.

Google scholar citations

<http://scholar.google.com.au/citations?user=HtiTsgwAAAAJ&hl=en>

Research Supervision

Current Students

1. He Zhao, Ph.D student at Monash University
Research project: Bayesian models for discrete data analysis with rich side information
When: Sep-2015 to now
Role: Principle supervisor
2. Sachith Seneviratne, Ph.D. student at Monash University
Research project: Automatic Generation of Statistical Models
When: Oct-2016 to now
Role: co-supervisor

Students Completed

1. Yusheng Chen, Master by research student at Monash
Research project: Multi-relational Learning with Supervised Information
When: Feb-2017 to Nov-2017
Role: co-supervisor

Teaching Commitment

Chief Examiner for the following units

FIT5149 Applied data analysis
FIT5196 Data wrangling

Teaching

Semester 2, 2018: FIT5149 Applied data analysis
Semester 1, 2018: FIT5196 Data wrangling
Semester 1, 2017: FIT5196 Data wrangling
Semester 2, 2016: FIT5149 Applied data analysis
Semester 2, 2016: FIT5196 Data wrangling
Online Teaching Period 3, 2016: FIT5196 Data wrangling

Professional Activities

Conference Programme Committee member:

2018:

The International Conference on Machine Learning (ICML)
Neural Information Processing Systems (NIPS)
The International Conference on Artificial Intelligence and Statistics (AISTATS)
the International Joint Conference on Artificial Intelligence (IJCAI)
The Association for the Advancement of Artificial Intelligence (AAAI)
The Conference on Uncertainty in Artificial Intelligence (UAI)
The Annual Meeting of the Association for Computational Linguistics (ACL)
The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)
Empirical Methods in Natural Language Processing (EMNLP)
The International Conference on Computational Linguistics (COLING)

2017

The Annual Meeting of the Association for Computational Linguistics (ACL)
Empirical Methods in Natural Language Processing (EMNLP)
The International Joint Conference on Natural Language Processing (IJCNLP)
The conference of the European Chapter of the Association for Computational Linguistics (EACL)
Neural Information Processing Systems (NIPS)
The International Conference on Artificial Intelligence and Statistics (AISTATS)
The Conference on Uncertainty in Artificial Intelligence (UAI)
The Association for the Advancement of Artificial Intelligence (AAAI)

2016

The Annual Meeting of the Association for Computational Linguistics (ACL)
The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)
The International Conference on Computational Linguistics (CoLing)
Neural Information Processing Systems (NIPS)
The International Conference on Machine Learning (ICML)
The International Conference on Artificial Intelligence and Statistics (AISTATS)
The Association for the Advancement of Artificial Intelligence (AAAI)

2015

The Annual Meeting of the Association for Computational Linguistics (ACL)
The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)
Empirical Methods in Natural Language Processing (EMNLP)
The 2014 Australasian Language Technology Association Workshop (ALTA)
Neural Information Processing Systems (NIPS)
The International Conference on Artificial Intelligence and Statistics (AISTATS)
The Asian Conference on Machine Learning (ACML)
The Australian Joint Conference on Artificial Intelligence (AI)

2014

The Annual Meeting of the Association for Computational Linguistics (ACL)
Empirical Methods in Natural Language Processing (EMNLP)
The 2014 Australasian Language Technology Association Workshop (ALTA)
The International Conference on Artificial Intelligence and Statistics (AISTATS)
The Asian Conference on Machine Learning (ACML)

2013

The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)
The International Joint Conference on Artificial Intelligence (IJCAI)

2011

The International Joint Conference on Artificial Intelligence (IJCAI)

2010

The Asian Conference on Machine Learning (ACML)

Journal reviewer:

TKDE

ACM Computing Surveys

Journal of Machine Learning Research

Machine Learning

IEEE Transactions on Neural Networks and Learning Systems

ACM transactions on Intelligent Systems and Technology

BMC Bioinformatics

Information Processing & Management

Neurocomputing