

Bo HAN

CONTACT

<http://tq0100r.github.io/>

TQ0100R@gmail.com

WORK EXPERIENCE

I have 10 years experience in Natural Language Processing and Social Media Analytics. My computational thinking mindset and fast prototyping skills help to achieve business success and my experience in start-ups and industry giants qualifies me from a full stack architect to an applied NLP consultant.

RedMarker, Sydney, Australia

Senior NLP Data Scientist

May 2017– Present

I am leading the optimisation of a compliance checking system [Artemis](#). My core responsibility involves applying NLP and active learning to improve system efficiency and accuracy.

Hugo AI, Sydney, Australia

Senior Data Scientist/NLP Engineer

March 2016– March 2017

I am leading the design and development work for automatic user profile summarisation which has become a key component in [production](#). The work features data crawling, cleaning, reconciliation and natural language generation from various social media sources. The deliverables facilitate the networking management for people like recruiters and business development managers.

IBM Research Australia, Melbourne, Australia

Research Scientist

April 2014– March 2016

I am responsible for developing and evaluating social media text analytics, including a text-based [geotagger](#), [Social Media Analytics in Real Time](#) and a biomedical [named entity recogniser](#). These text analytics combine advanced NLP techniques in an adapted big data processing framework. The deliverables have been successfully applied in many internal research prototypes and client projects.

The University of Melbourne/NICTA, Melbourne, Australia

PhD Researcher

July 2010 – March 2014

I am working on applying NLP/ML methods to improve the utility of social media for downstream text mining applications.

Research Assistant

July 2011 – January 2012

I am a contributor for an [IBES](#) project: “Early detection and mitigation of disastrous events with broadband-enabled social interaction technologies”.

Microsoft Research Asia, Beijing, China

Full-time Research Intern

September 2009– June 2010

I helped to build an English writing assistant prototype which has been shipped as the dictionary tab in [Bing Search](#) in China. My responsibility includes bilingual text mining, short sentence categorisation, incremental bilingual text crawling.

Full-time Research Intern

January 2007 – August 2007

I helped to improve web corpus quality for various research projects and built a internal question and answer community.

PROFESSIONAL SKILLS

- Frontend: HTML/CSS/Javascript including JQuery
- Languages: Python (Proficient), R, bash scripts, Java, C#, C/C++
- Modelling: scikit-learn, spaCy, word2vec/fasttext, TextBlob, Pattern, Keras (TensorFlow)
- Backend: popular SQL and non-SQL databases e.g., MySQL, Postgres, MongoDB, Redis
- Infrastructure: DevOps (Git, Jenkins, Ansible), AWS (EC2 and S3), clusters (PBS), Vagrant

EDUCATION

The University of Melbourne, Melbourne, Australia

Doctor of Philosophy

July 2010 – August 2014

Dalian University of Technology, Dalian, China

Bachelor and Master of Engineering

September 2003 – December 2009

PUBLICATIONS & PATENTS

I am actively involved in professional activities including publishing research and demo papers; giving invited talks in universities (Sydney University, RMIT, Tianjin University, Saarland University, TU Darmstadt) and industrial institutions (DFKI); being reviewers in top-tier conferences and journals (e.g., ACL, EMNLP, NAACL, COLING, NLE); co-organising two international workshops on social media analytics ([W-NUT 2015](#) and [W-NUT 2016](#)). I maintain a full list of my publications in [Google Scholar](#) and I sometimes also open sourced my projects on [GitHub](#).

US Patent: Dynamic modelling of geospatial words in social media (US 9405743 B1)

US Patent: System and method for unstructured metadata correlation (US YOR920160975US1)

Bo Han, Will Radford, Anas Cadilhac, Art Harol, Andrew Chisholm, Ben Hachey, “Post-edit Analysis of Collective Biography Generation”, in WWW 2017, Perth, Australia.

Bo Han, Afshin Rahimi, Leon Derczynski and Timothy Baldwin, “Twitter Geolocation Prediction Shared Task”, in 2nd W-NUT 2016, 213–217, Osaka, Japan

Bo Han and Antonio Jimeno Yepes and Andrew MacKinlay and Lianhua Chi, “Temporal Modelling of Geospatial Words in Twitter”, in ALTA 2016, 144–148, Melbourne, Australia

Bo Han, Paul Cook and Timothy Baldwin, “Text-based Twitter User Geolocation Prediction”, In *Journal of Artificial Intelligence Research*, 49, pp. 451–500, 2014. **(100+ citations)**

Spandana Gella, Paul Cook and Bo Han, “Unsupervised Word Usage Similarity in Social Media Texts”, in *SEM2013, 248–253, Atlanta, GA, USA. **(Best short paper)**

Bo Han, Paul Cook and Timothy Baldwin, “A Stacking-based Approach to Twitter User Geolocation Prediction”, In *ACL 2013*, Demo Session, pp. 7–12, Sofia, Bulgaria, 2013.

Bo Han, Paul Cook and Timothy Baldwin, “Lexical Normalisation of Short Text Messages”, In *ACM Transactions on Intelligent Systems and Technology* 4(1), pp. 5:1–5:27, 2013.

Bo Han, Paul Cook and Timothy Baldwin, “Automatically Constructing a Normalisation Dictionary for Microblogs”, In *Proceedings of EMNLP 2012*, 421–432, Jeju, South Korea, 2012. **(100+ citations)**

Bo Han and Timothy Baldwin, “Lexical normalisation of short text messages: Makn sens a #twitter”, In *ACL 2011*, 368–378, Portland, OR, USA, 2011. **(350+ citations)**

Xiaohua Liu, Bo Han, Kuan Li, Stephan Hyeonjun Stiller, and Ming Zhou, “SRL-based verb selection for ESL”, In *EMNLP 2010*, 1068–1076, Cambridge, MA, USA, 2010.

REFEREES

Timothy Baldwin

Professor/ARC future fellow
Melbourne University

Ming Zhou

Vice President
Microsoft Research Asia

Jennifer Lai

Executive Architect
IBM Research/Watson