Bo HAN

CONTACT http://tq010or.github.io/ TQ010OR@gmail.com

WORK
EXPERIENCE

I have 11 years research and project development experience in the realm of media, Internet services, finance and healthcare. My computational thinking mindset, prototyping skills and product experience enable me to identify data-driven business opportunities and deliver impact-oriented solutions in both start-up and enterprise environments.

Kaplan   Sydney, Australia

Advanced NLP Specialist   05/2017 – Present

I am optimising a ML-based compliance system Artemis in financial sector. My work consists of three parts: apply active learning and other pragmatic methods to improve data annotation; boost system speed and scalability with mature cloud computing services; co-develop an in-house textual data anonymisation system to meet the privacy protection regulations.

Hugo AI   Sydney, Australia

Senior Data Scientist   03/2016 – 03/2017

I lead the design and development for automatic user profile summarisation. The work features data crawling, cleaning, reconciliation and natural language generation from various social media sources. The deliverables facilitate online recruiters for more effective and efficient talent acquisitions.

IBM Research Australia   Melbourne, Australia

Research Scientist   04/2014 – 03/2016

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 aim to enhance IBM Watson in dealing with domain specific tasks. They combines advanced NLP techniques in an adapted big data processing framework. The deliverables have been successfully applied in many client engaged projects such as Australian Open.

University of Melbourne/NICTA   Melbourne, Australia

PhD Researcher   07/2010 – 03/2014

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

Research Assistant   07/2011 – 01/2012

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   09/2009 – 06/2010

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   01/2007 – 08/2007

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

EDUCATION   **University of Melbourne**   Melbourne, Australia

Doctor of Philosophy   07/2010 – 04/2014

Dalian University of Technology   Dalian, China

Bachelor and Master of Engineering   09/2003 – 12/2009

PROFESSIONAL SKILLS	

<table>
<tr><td>PROFESSIONAL
SKILLS</td><td>

- Frontend: HTML, CSS, Javascript (e.g., JQuery)
- Languages: Python, R, bash scripts, Java, C#, C/C++
- Modelling: scikit-learn, spaCy, word2vec/fasttext, Keras (TensorFlow)
- Backend: MySQL, Postgres, MongoDB, Redis, Memcached
- Cloud: popular AWS services such as EC2, Lambda, S3 and CloudWatch
- Infrastructure: JIRA, Git, Jenkins/Bamboo, Ansible

</td></tr>

<tr><td>PUBLICATIONS &
PATENTS</td><td>

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, CMCRC); 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)

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, Paul Cook and Timothy Baldwin, "Text-based Twitter User Geolocation Prediction", In *Journal of Artificial Intelligence Research*, 49, pp. 451–500, 2014. **(150+ 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, "Geolocation Prediction in Social Media Data by Finding Location Indicative Words", In *COLING 2012*, 1045–1062, Mumbai, India.

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. **(130+ 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. **(400+ 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.

</td></tr>

<tr><td>REFEREES</td><td>

Timothy Baldwin
Professor/ARC future fellow
University of Melbourne

Ming Zhou
Vice President
Microsoft Research Asia

Jennifer Lai
Executive Architect
IBM Research/Watson

</td></tr>
</table>