# Bo HAN

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

## WORK EXPERIENCE

**Hugo AI**, Sydney, Australia

*Senior Data Scientist*                                                                                    **March 2016– Present**

I am leading design and development work for automatic text summarisation for a person or an organisation from their various social media sources. The deliverable has become a key asset in production. The work features social data crawling, cleaning, reconciliation and natural language generation.

**IBM Research Australia**, Melbourne, Australia

*Research Scientist*                                                                                    **April 2014– March 2016**

I am responsible for developing and evaluating text analytics, including a text-based geotagger, Social Media Analytics in Real Time and a biomedical named entity recogniser. The deliverables have been successfully applied in both internal research prototypes and external client projects.

**The University of Melbourne**, Melbourne, Australia

*Teaching Assistant* for COMP90049 Knowledge Technologies                        **July 2013– October 2013**

*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.

## EDUCATION

**The University of Melbourne**, Melbourne, Victoria, Australia

*Doctor of Philosophy*                                                                            **July 2010 – August 2014**

**Dalian University of Technology**, Dalian, Liaoning, China

*Master of Engineering*                                                                        **September 2007 – December 2009**

*Bachelor of Engineering*                                                                        **September 2003 – July 2007**

## PROFESSIONAL SKILLS

I have a strong background in Natural Language Processing and Social Media Analytics. My computational thinking and fast prototyping help to achieve business success and my experience in start-ups and industry giants qualify me from a full stack architect to a team leader.

- Frontend: HTML and Javascript including JQuery
- Programming: Proficient in Python, bash scripts and GNU utilities
- Modelling: Pandas, scikit-learn, NLP suites (e.g., spaCy, gensim, Pattern, TextBlob, nltk), R (RStudio) and Deep Learning (Keras with TensorFlow or Theano)
- Backend: popular SQL and non-SQL databases e.g., MySQL, Postgres, MongoDB
- Infrastructure and services: DevOps (Git and Jenkins), AWS (EC2 and S3), computing clusters (PBS system)

| PROFESSIONAL ACTIVITIES | I am involved in a range of professional activities including giving invited talks in universities and industrial institutions, being reviewers in top-tier conferences (e.g., ACL, EMNLP, NAACL, COLING, NLE), co-organising international workshops on social media analytics. |
|---|---|

Co-organiser for 2nd workshop on Noisy User-generated Text (2nd W-NUT)
Co-founder and co-organiser for 1st workshop on Noisy User-generated Text (1st W-NUT)
Co-organiser for English Tweet Text Geotagging Shared Task in Osaka, Japan
Co-organiser for English Lexical Normalisation Shared Task in Beijing, China
I sometimes open sourced my projects on GitHub.

Presentations and invited talks:
12/2016: Shared Task of English Geotagging, Osaka, Japan.
05/2016: Guest lecture on Geotagging, Sydney University, Australia.
10/2015: Introduction to Artificial Intelligence Apps, The University of Melbourne, Australia.
08/2015: Introduction to Geotagging, IBM CRL, China.
08/2015: Introduction to Geotagging, Tianjin University, China.
07/2015: Shared Task of English Lexical Normalisation, Beijing, China.
07/2015: Investigating Public Health Surveillance using Twitter, Beijing, China.
11/2014: Social media text processing, RMIT, Australia.
08/2013, Improving the Utility of NLP on Social Media, DFKI/COLI, Saarbrucken, Germany.
08/2013, A Stacking-based Approach to Twitter User Geolocation Prediction, Sofia, Bulgaria.
08/2013, Improving the Utility of NLP on Social Media, TU-Darmstadt, Germany.
08/2012, Event Detection over Social Media, Kyoto University, Japan.
07/2012, Automatically Constructing a Normalisation Dictionary for Microblogs, Jeju, Korea.
07/2011, Lexical Normalisation for Twitter Messages, Portland, USA.

| SELECTED PUBLICATIONS | I maintain a full list of my publications in Google Scholar. |
|---|---|

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

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 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.

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-CoNLL 2012*, 421–432, Jeju, Republic of Korea, 2012.

Bo Han and Timothy Baldwin, "Lexical normalisation of short text messages: Makn sens a #twitter", In *ACL 2011*, 368–378, Portland, OR, USA, 2011. **(300+ 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.