Bo HAN

CONTACT

Address: WS 19.071, Southbank, VIC, Australia   Office: +61-3-8646-5545
Website: http://tq010or.github.io/   Email: bohan.ibm@au1.ibm.com

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**

WORK EXPERIENCE

IBM Research, Melbourne, Victoria, Australia

Postdoctoral researcher   **April 2014– To date**

Responsible for developing and evaluating social media analytics, big data processing and analysis

The University of Melbourne, Melbourne, Victoria, Australia

Teaching Assistant   **July 2013– October 2013**

COMP90049 Knowledge Technologies

Research Assistant   **July 2011 – January 2012**
IBES project for "Early detection and mitigation of disastrous events with broadband-enabled social interaction technologies"

Microsoft Research Asia, Beijing, China

Fulltime Research Intern   **September 2009– June 2010**

Bilingual short text mining, English as a Second Language learner writing assistance, short text classification, Semantic Role Labelling for news tweets, incremental web page crawling.

Fulltime Research Intern   **January 2007 – August 2007**
Web data cleansing, Web 2.0 design and implementation, bilingual short text mining.

SKILLS

Fast prototyping and strong problem solving skills
Professional communication and presentation skills
Big data processing and analysis (R, python, Unix/Linux scripting, PySpark, computing clusters)
Full stack programmer (from Jquery to MongoDB)

RESEARCH

Natural Language Processing, Social Media Analytics in Real Time, Big Data Consulting

PROFESSIONAL ACTIVITIES

Co-organiser for an ACL 2015 workshop on Noisy User-generated Text (W-NUT)
Co-organiser for English Lexical Normalisation Shared Task

Academic visits and Invited Talks:
20/10/2015: Introduction to Artificial Intelligence Apps, The University of Melbourne, Australia.
04/08/2015: Introduction to Geotagging, IBM CRL, China.
03/08/2015: Introduction to Geotagging, Tianjin University, China.
31/07/2015: Shared Task of English Lexical Normalisation, Beijing, China.
30/07/2015: Investigating Public Health Surveillance using Twitter, Beijing, China.
12/11/2014: Social media text processing, RMIT, Australia.
13/08/2013, Improving the Utility of NLP on Social Media, DFKI/COLI, Saarbrucken, Germany.
05/08/2013, A Stacking-based Approach to Twitter User Geolocation Prediction, Sofia, Bulgaria.
01/08/2013, Improving the Utility of NLP on Social Media, TU-Darmstadt, Germany.
28/08/2012, Event Detection over Social Media, Kyoto University, Japan.
12/07/2012, Automatically Constructing a Normalisation Dictionary for Microblogs, Jeju, Korea.
21/07/2011, Lexical Normalisation for Twitter Messages, Portland, USA.

Reviewers: ACL 2014, 2015; EMNLP 2013, 2014, 2015, Natural Language Engineering

CODEBASE AND DEMOS	Twitter user geolocation prediction codebase: https://github.com/tq010or/acl2013 Locations Indicative Words per city: http://goo.gl/2XHtFc Twitter text normalisation lexicon: http://goo.gl/Tqn0O3
PATENT	US Patent: Dynamic modelling of geospatial words in social media (filed)
SELECTED PUBLICATIONS	Baldwin, Timothy, Marie-Catherine de Marneffe, Bo Han, Young-Bum Kim, Alan Ritter and Wei Xu, "Shared Tasks of the 2015 Workshop on Noisy User-generated Text: Twitter Lexical Normalization and Named Entity Recognition", in *Proceedings of the ACL 2015 Workshop on Noisy User-generated Text (W-NUT)*, pp. 126–135, Beijing, China, 2015. Antonio Jimeno Yepes, Andrew MacKinlay, Bo Han, "Investigating Public Health Surveillance using Twitter", in *BioNLP 2015*, 157–162, Beijing, China, 2015. Antonio Jimeno Yepes, Andrew MacKinlay, Bo Han, "Identifying Diseases, Drugs and Symptoms in Twitter", in *MedInfo 2015*, 157–162, Sao Paulo, Brazil, 2015. Bo Han, Antonio Jimeno Yepes, Andrew MacKinlay, Qiang Chen, "Identifying Twitter Location Mentions", in *ALTA 2014*, shared task session, pp. 157–162, Melbourne, Australia. Paul Cook, Bo Han and Timothy Baldwin, "Statistical Methods for Identifying Local Dialectal Terms from Corpora of GPS-tagged Documents", In *Dictionaries*, 35, pp. 248–271, 2014. 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, 2013. (**Best short paper award**) Bo Han, Paul Cook and Timothy Baldwin, "unimelb: Spanish Text Normalisation", In *Proceedings of Tweet-norm: Tweet Normalization Workshop at SEPLN 2013*, 67–71, Madrid, Spain, 2013. 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. Xianchao Zhang, Wenxin Liang, Shaoping Zhu, Bo Han, "Automatic seed set expansion for trust propagation based anti-spam algorithms", In *Information Sciences*, 232, 167–187, Elsevier, 2013. Yang Liao, Masud Moshtaghi, Bo Han, Shanika Karunasekera, Ramamohanarao Kotagiri, Timothy Baldwin, Aaron Harwood and Philippa Pattison, "Mining Micro-Blogs: Opportunities and Challenges", In *Social Networks: Computational Aspects and Mining*, Ajith Abraham and Aboul Ella Hassanien (eds.), Springer: London, UK, pp. 129–159. 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. Timothy Baldwin, Paul Cook, Bo Han, Aaron Harwood, Shanika Karunasekera and Masud Moshtaghi, "A Support Platform for Event Detection using Social Intelligence", In *EACL 2012*, Demo Session, 69–72, Avignon, France, 2012. Bo Han, Marco Lui and Timothy Baldwin, "Melbourne Language Group Microblog Track Report", In *TREC 2011*, NIST Special Publication: SP 500–295. Bo Han and Timothy Baldwin, "Lexical normalisation of short text messages: Makn sens a #twitter", In *ACL 2011*, 368–378, Portland, OR, USA, 2011. (**200+ citations in 4 years.**) Xiaohua Liu, Bo Han, and Ming Zhou, "Correcting verb selection errors for ESL with the perceptron", In *CICLing 2011*, 411–423, Tokyo, Japan, 2011. 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. Xiaohua Liu, Kuan Li, Bo Han, Ming Zhou, Long Jiang, Zhongyang Xiong, and Changning Huang, "Semantic role labeling for news tweets", In *Coling 2010*, 698–706, Beijing, China, 2010. Xiaohua Liu, Kuan Li, Bo Han, Ming Zhou, Long Jiang, Daniel Tse, Zhongyang Xiong, "Collective Semantic Role Labeling on Open News Corpus by Leveraging Redundancy", In *Coling* 2010, 725–729, Beijing, China, 2010. Xianchao Zhang, Bo Han, and Wenxin Liang, "Automatic seed set expansion for trust propagation based anti-spamming algorithms", In *WIDM 2009*, 31–38, Hongkong, China, 2009.

REFEREES

Timothy Baldwin
Professor, ARC future fellow
Melbourne University
e-mail: *tb@ldwin.net*

Ming Zhou
Principal Researcher
Microsoft Research Asia
e-mail: *mingzhou@microsoft.com*

Jennifer Lai
Senior Manager
IBM Research
e-mail: *jlai@us.ibm.com*