

Zhongyu WEI

CONTACT INFORMATION	Room 711, M.M.W. Engineering Building The Chinese University of Hong Kong Shatin, Hong Kong, China Website: http://www.se.cuhk.edu.hk/~zywei	Tel: +(852) 6841-1735 Fax: +(852) 2603-5505 E-mail: zywei@se.cuhk.edu.hk
RESEARCH INTERESTS	Cross-media summarization, social media analysis, microblog search, natural language processing	
EDUCATION	The Chinese University of Hong Kong (CUHK) , Hong Kong, China Ph.D. Candidate, Department of Systems Engineering and Engineering Management (SEEM), (August 2010 - Now) <ul style="list-style-type: none">• Adviser: Professor Kam-Fai Wong• Area of Study: Social Computing, Information Retrieval, Natural Language Processing• Estimated Graduation Time: September 2014 Harbin Institute of Technology (HIT) , Harbin, China M.Eng., Intelligence Computing Research Centre, (August 2008 - July 2010) B.Eng., Department of Computer Science and Technology, (August 2004 - July 2008)	
SELECTED PUBLICATIONS	<ol style="list-style-type: none">1. Gaoyan Ou, Wei Chen, Dongqing Yang, Zhongyu Wei, Binyang Li, Tengjiao Wang and Kam-Fai Wong, <i>Exploiting Community Emotion for Microblog Event Detection</i>, The Conference on Empirical Methods in Natural Language Processing (EMNLP), October, 2014, accepted.2. Zhongyu Wei, Wei Gao, <i>Utilizing Microblogs for Automatic News Highlights Extraction</i>, The 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland, August, 2014.3. Zhongyu Wei, Wei Gao, <i>Ranking Model Selection and Fusion for Effective Microblog Search</i>, SIGIR 2014 Workshop on Social Media Retrieval and Analysis (SoMeRA), Gold Coast, Australia, July, 2014.4. Binyang Li, Lanjun Zhou, Zhongyu Wei, Kam-Fai Wong, Ruifeng Xu and Yunqing Xia, <i>Web Information Mining and Decision Support Platform for the Modern Service Industry</i>, The 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Demo Paper, Baltimore, USA, June 22-27, 2014.5. Lanjun Zhou, Binyang Li, Zhongyu Wei and Kam-Fai Wong, <i>An Open Discourse Corpus for Chinese with Annotated Explicit Discourse Connectives</i>, The 9th edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May, 2014.6. Tarek El-Ganainy, Zhongyu Wei, Walid Magdy, and Wei Gao, <i>QCRI at TREC 2013 Microblog Track</i>, TREC2013 (2nd place among 65 automatic runs).7. Zhongyu Wei, Junwen Chen, Wei Gao, Binyang Li, Lanjun Zhou, Yulan He and Kam-Fai Wong, <i>An Empirical Study on Uncertainty Identification in Social Media Context</i>, The 51st Annual Meeting of the Association for Computational Linguistics (ACL), Short Paper, Sofia, Bulgaria, August, 2013.8. Zhongyu Wei, Yulan He, Wei Gao, Lanjun Zhou, Binyang Li and Kam-Fai Wong, <i>Mainstream Media Behavior Analysis on Twitter: A Case Study on UK General Election</i>, The 24th ACM Conference on Hypertext and Social Media (HyperText), Paris, France, May, 2013.	

9. **Zhongyu Wei**, Yulan He, Simon Buckingham Shum, Rebecca Ferguson, Wei Gao and Kam-Fai Wong, *A Self-Training Framework for Automatic Identification of Exploratory Dialogue*, The 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), Samos, Greece, March, 2013.
10. Rebecca Ferguson, **Zhongyu Wei**, Yulan He, Simon Buckingham Shum, *An Evaluation of Learning Analytics to Identify Exploratory Dialogue in Online Discussions*, The 3rd ACM International Conference on Learning Analytics and Knowledge (LAK), Leuven, Belgium, April, 2013.
11. Wei Gao, **Zhongyu Wei** and Kam-Fai Wong, *Microblog Search and Filtering with Time Sensitive Feedback and Thresholding Based on BM25*, TREC2012 (Microblog track).
12. **Zhongyu Wei**, Wei Gao, Lanjun Zhou, Binyang Li, and Kam-Fai Wong, *Exploring Tweets Normalization and Query Time Sensitivity for Twitter Search*, TREC2011 (Microblog track).
13. Lanjun Zhou, Binyang Li, Wei Gao, **Zhongyu Wei** and Kam-Fai Wong, *Unsupervised Discovery of Discourse Relations for Eliminating Intra-sentence Polarity Ambiguities*, The Conference on Empirical Methods in Natural Language Processing (EMNLP), July, 2011.
14. **Zhongyu Wei**, Jun Xu, Xiaolong Wang. *One-class Classification based Finance News Story Recommendation*, Journal of Computational Information Systems, 2009, 5(6): 1625-1631.

PROFESSIONAL EXPERIENCE

Research Intern, Qatar Computing Research Institute(QCRI), Qatar, (June 2013 - December 2013)

- Mentor: Dr. Wei Gao
- Project: Utilizing Microblogs for Automatic News Highlights Extraction

Research Intern, Knowledge Media Institute (KMi), The Open University, UK, (April 2012 - June 2012)

- Mentor: Dr. Yulan He and Prof. Simon Buckingham Shum
- Project: Exploratory Discourse Detection

CONFERENCE REVIEW

Secondary reviewer during the PhD study

2014: SigIR, COLING, EMNLP
 2013: ACL, EMNLP, IJCAI, RANLP, AIRS
 2012: ACL, EMNLP, PACLIC, AIRS
 2011: EMNLP

PROJECT EXPERIENCE

Utilizing Microblogs for Automatic News Highlights Extraction

Researcher, June 2013 - December 2013, Arabic Language Technologies, Qatar Computing Research Institute, Qatar

This project aims to generate sentences highlights for readers to quickly access the gist of a given news article. We propose a novel method to improve news highlights extraction by using microblogs. Given a news article, we formulate the problem as two extraction tasks with the help of cross-media correlation information, sentences extraction from news article and tweets extraction from tweet set. We proposed a supervised and an unsupervised method for solving both tasks. We also constructed two corpus for evaluation. Experiment results showed that our supervised methods outperformed several strong baselines and the unsupervised method can perform similar to the supervised one.

Emergent Rumor Detection and Credibility Ranking

Researcher, September 2012 - December 2014, Web Intelligence System Engineering Center, CUHK, Hong Kong

This project aims to identify rumor from social media and rank rumors in terms of harmfulness. We propose a novel framework to spot rumor event from three aspects: (1) text uncertainty, (2) opinion controversy, (3) opinion based credibility. We will deliver an emergent rumor detection system based on the framework.

Web Information Mining and Decision Support Platform for the Modern Service Industry

Project coordinator, September 2010 - August 2012, Web Intelligence System Engineering Center, CUHK, Hong Kong

This project (MODEST) aims to provide enterprises with the services of retrieving news from websites, extracting commercial information, exploring customers' opinions, and analyzing collaborative/competitive social networks. In this way, enterprises can improve the competitive abilities and facilitate potential collaboration activities. Currently, MODEST is applied to the pillar industries of Hong Kong, including innovative finance, modern logistics, information technology, etc. [\[DEMO\(in Chinese\)\]](#)

Exploratory Discourse Detection

Researcher, April 2012 - June 2012, KMi, The Open University, United Kingdom

This project aims to identify exploratory dialogues from online learning material. Exploratory dialogue is desirable in online learning environments, since training learners to use it has been shown to improve learning outcomes. We propose a self-training framework to identify exploratory dialogue and apply it to [Social Learn Platform](#) to assist distance education.

Mainstream Media Behavior Analysis on Twitter

Researcher, September 2011 - February 2012, Web Intelligence System Engineering Center, CUHK, Hong Kong

This project aims to investigate the behavior of mainstream media on Twitter and study how they exert their influence to shape public opinion during the UK's 2010 General Election. We first propose an empirical measure to quantify mainstream media bias based on sentiment analysis and show that it correlates better with the actual political bias in the UK media than the pure quantitative measures. We then compare the information diffusion patterns from different categories of sources and reveal that the role of mainstream media in shaping public opinion is being challenged by journalists.

Microblog Search at TREC

TREC2013: Our system focuses on search of tweets and re-ranking of search results which utilizes concurrent Web search results for pseudo relevance feedback (PRF) and the ensemble of multiple rankers for re-ranking retrieved tweets. We conduct extensive studies on applying learning to rank for microblog search re-ranking, consisting of feature engineering, validation dataset selection and combining various ranking models. For query expansion, we expand the queries by making use of Google search results as external resource in addition to PRF only using in-collection search results. **(2nd place among 65 automatic runs)**

TREC2012: Our system is built based on the traditional BM25 relevance model, in which specific techniques are tried out to respond to the need of finding relevant tweets. We apply a peak detection algorithm for the process of blind feedback. We also try to automatically combine the search results of multiple retrieval techniques.

TREC2011: Microblog texts like tweets are generally characterized by the inclusion of a large proportion of irregular expressions, such as ill-formed words, which can lead to significant mismatch between query terms and tweets. In addition, Twitter queries are distinguished from Web queries with many unique characteristics, one of which reflects the clearly distinct temporal aspects of Twitter search behavior. We incorporate tweets normalization and time-sensitive query detection to explore two obvious characteristics of search in microblog contexts, i.e., the ill-formed tweet words and the time-sensitivity of queries.

Financial News Ranking Module in an Intelligence Information Search Engine

Researcher, May 2009 - July 2010, Intelligence Computing Research Center, HIT, China

This project aims to rank the financial news according to its importance to users based on individual stock. We propose several novel features from financial news and combine all of them by learning to rank model for financial news ranking. The developed ranking model has been adopted in the intelligence information search engine effectively.

PROFESSIONAL SKILLS

General: Fast modeling and prototyping; Data analysis; Familiar with Twitter data processing
Programming: Java, Python, Unix/Linux shell scripting, R, C/C++

TEACHING EXPERIENCE

Teaching Asistant

Engineering Entrepreneurship, SEEM, CUHK, 2011, 2012, 2013, 2014
 Information Systems Analysis and Design, SEEM, CUHK, 2012
 Information Technology Management, SEEM, CUHK, 2012, 2013
 Information Systems Management, SEEM, CUHK, 2011

AWARDS

July 2010	Outstanding Graduates in Province (top 3%)
July 2010	Outstanding Postgraduate Students in HIT (top 1%)
July 2008	Outstanding Graduates in Province (top 3%)
September 2006	Outstanding Student Leaders in Province (top 1%)

REFEREES

Professor Kam-Fai WONG

PhD, Professor
 The Chinese University of Hong Kong
 Email: kfwong@se.cuhk.edu.hk

Dr. Wei GAO

PhD, Scientist
 Qatar Computing Research Institute
 Email: wgao@qf.org.qa

Dr. Yulan HE

PhD, Senior lecturer
 Aston University
 Email: y.he9@aston.ac.uk