Tao CHEN

Mobile: (+65) 9116-8158
Email: taochen@comp.nus.edu.sg
http://www.comp.nus.edu.sg/~taochen

AS6 04-13, COM 1,
13 Computing Drive,
Singapore, 117417

Research Interest

My research lies at the intersection of social media, multimedia, natural language processing and machine learning (e.g., topic modeling and recommender system). I study both the user generated content and user behaviors in social media platforms.

Education

National University of Singapore, Singapore — Aug 2010 - Current
- Ph.D in Computer Science, GPA 4.69/5.0
- Advisor: Associate Professor Min-Yen KAN

East China Normal University, Shanghai, China — Sep 2006 - Jun 2010
- Bachelor in Software Engineering, GPA 3.73/4.0, Rank 2/161
- Awarded School Excellent Thesis and Shanghai Excellent Graduate

Publications

Xiangnan He, **Tao Chen**, Min-Yen Kan and Xiao Chen (2015). Review-aware Explainable Recommendation by Modeling Aspects. *In Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM'15).*

Bang Hui Lim, Dongyuan Lu, **Tao Chen** and Min-Yen Kan (2015). #mytweet via Instagram: Exploring User Behaviour across Multiple Social Networks. *In Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'15).*

Tao Chen, Naijia Zheng, Yue Zhao, Muthu Chandrasekaran and Min-Yen Kan (2015). Interactive Second Language Learning from News Websites. *In Proceedings of ACL Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA'15).*

Tao Chen, Hany Salaheldeen, Xiangnan He, Min-Yen Kan and Dongyuan Lu (2015). VELDA: Relating an Image Tweet's Text and Images. *In Proceedings of the 29st AAAI Conference on Artificial Intelligence (AAAI'15).*

Tao Chen, Dongyuan Lu, Min-Yen Kan and Peng Cui (2013). Understanding and Classifying Image Tweets. *In Proceedings of the 21st ACM Multimedia (MM'13).*

Tao Chen and Min-Yen Kan (2013). Creating a Live, Public Short Message Service Corpus: The NUS SMS Corpus. *Language Resources and Evaluation*, 47(2)(2013).

Aobo Wang, **Tao Chen** and Min-Yen Kan (2012). Re-tweeting from a linguistic perspective. *In Proceedings of the NAACL-HLT 2012 Workshop on Language in Social Media (LSM'12).*

Research Projects

1. Social Multimedia:
   - **Understanding and Classifying Image Tweet** [MM'13]
     We study a large corpus of image tweets in order to uncover what people post about and the correlation between the tweet's image and its text. We show that an important functional distinction is between visually-relevant and visually-

Last updated: Nov 11, 2015

irrelevant tweets, and that we can successfully build an automated classifier utilizing text, image and social context features to distinguish these two classes.

- **VELDA: Relating an Image Tweet's Text and Images** [AAAI'15]

  Image tweets are becoming a prevalent form of social media, but little is known about their content - textual and visual - and the relationship between the two mediums. Our analysis of image tweets shows that while visual elements certainly play a large role in image-text relationships, other factors such as emotional elements, also factor into the relationship. We develop Visual-Emotional LDA (VELDA), a novel topic model to capture the image-text correlation from multiple perspectives (namely, visual and emotional).

- **Personalized Image Tweets Recommendation**

  Twitter displays tweets in a chronological order. However, the information overload problem troubles users, as users are only interested in a small portion of tweets. To address this, we re-rank image tweets according to user's preference. We leverage the latent factor model, a typical collaborative filtering technique, to model user's latent interests in image tweets.

2. Social Media:

- **Re-tweeting from a Linguistic Perspective** [LSM'12]

  We study the content of tweets to uncover linguistic tendencies of shared microblog posts (re-tweets), by examining surface linguistic features, deeper parse-based features and Twitter-specific conventions in tweet content. We show how these features correlate with a functional classification of tweets, thereby categorizing peoples writing styles based on their different intentions on Twitter. We find that both linguistic features and functional classification contribute to re-tweeting.

- **Exploring User Behaviour across Multiple Social Networks** [ASONAM'15]

  We study how users of multiple online social networks (OSNs) employ and share information by studying a common user pool that use 6 OSNs Flickr, Google+, Instagram, Tumblr, Twitter, and YouTube. We explore and analyse the factors that affect user behaviour. In particular, we take a temporal and topical approach in studying the sharing behaviour of users, and we show how users exhibit distinct patterns on different networks.

  This is an undergraduate final year project that I supervised as a PhD mentor.

3. Natural Language Processing:

- **Interactive Second Language Learning from News Websites** [NLP-TEA'15]

  We propose a web browser extension that allows readers to learn a second language vocabulary while reading news online. Injected tooltips allow readers to look up selected vocabulary and give interactive tests to assess vocabulary mastery. In particular, we propose techniques for word sense disambigution (WSD) to aid transation quality and constructing appropriate distractors for multiple-choice word mastery tests.

  This is a joint project with two final year undergraduates. This project is supported by Validating Startup Concept (VaSCo) Award (S$10K) from School of Computing, National University of Singapore.

- **Creating a Live, Public Short Message Service Corpus** [LRE'13]

  SMS message is a popular and personal form of communication. Due to the privacy issue, public SMS corpora are few, which hinders the related research. To address this, we collect a public English and Chinese SMS corpus via crowdsourcing, and pay particular attention to privacy issues. To date, the corpus consist

of 45,718 English SMS messages, and 31,465 Chinese SMS messages, which is the largest public SMS corpus in terms of both English and Chinese languages, independently.

Experience	**Research Assistant,** National University of Singapore	Sep 2014 - Current

- Advisor: Min-Yen KAN
- Working on second language learning from reading the news.

Software Engineering Intern, Google, New York　　　　Jun 2013 - Sep 2013
- Worked in review summarization team and focused on sentiment analysis internationalization.

Software Engineering Intern, IBM, Shanghai, China　　　　Jun 2009 - Sep 2009
- Developed an IE extension that could bookmark, comment and share web pages within IBM internal social network.

Teaching　　**Teaching Assistant,** National University of Singapore　　Semester 1 2013/2014
- Course: CS2106 Introduction to Operating System
- Led labs, answered students' questions and graded projects.

Teaching Assistant, East China Normal University　　Semester 1 2008/2009
- Course: Software Test and Quality Assurance
- Designed and implemented a prototype bug management system for this course and assisted the instructor in the lab sessions.

Student Mentoring　　Undergraduate Final Year Project, National University of Singapore
- Low Wee Heng: Citation Provenance (Spring 2012)
- Nhat Minh Le: Link Prediction (Fall 2013)
- Bang Hui Lim: Cross-Posting in Social Networks (Fall 2014)
- Yue Zhao and Naijia Zheng: Second Language Learning via News Websites (Fall 2014)

Undergraduate Research Opportunities Programme, National University of Singapore
- Low Wee Heng: Word Sense Disambiguation Using Decision Trees (Fall 2010)
- Tamisa Huangwongsri: Chinese Word Segmentation by Conditional Random Fields (Fall 2010)
- Eric Yulianto: Citation Function Classification (Spring 2012)

Services　　**Reviewer**: Information Retrieval Journal, Language Resources and Evaluation Journal, 2014 ACM International Workshop on Internet-Scale Multimedia Management, 2016 IEEE International Conference on Multimedia and Expo
Sub-Reviewer: 2011 International Conference on Asian Digital Libraries
Session Chair: The 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining
Volunteer: AAAI 2015, MLSS 2012

Awards & Scholarships	Validating Startup Concept (VaSCo) Award from School (S$10K)	Sep 2015
	AAAI 2015 Student Travel Scholarship	Jan 2015
	Google Anita Borg Memorial Scholarship: Asia Pacific	Sep 2014
	1st Place in Elsevier CodeForScience Singapore Competition	Sep 2012
	Machine Learning Summer School (MLSS) Singapore Scholarship	Jul 2011
	National University of Singapore Research Scholarship	2010 - 2014
	IBM Chinese Excellent Student Scholarship	Sep 2009

	National Inspirational Scholarship	Oct 2007
	First-class Scholarship in East China Normal University	2007 - 2010

Skills   Programming: Java, Python, C++, SQL, JSP, HTML, Javascript
Language: English (fluent), Chinese (native)

References   Dr. Min-Yen KAN, Associate Professor
School of Computing, National University of Singapore
Email: kanmy@comp.nus.edu.sg
Tel: (+65) 6516-1885

Dr. Chew Lim TAN, Professor
School of Computing, National University of Singapore
Email: tancl@comp.nus.edu.sg
Tel: (+65) 6516-2900

Dr. Hwee Tou NG, Professor
School of Computing, National University of Singapore
Email: nght@comp.nus.edu.sg
Tel: (+65) 6516-8951

Dr. Peng CUI, Assistant Professor
Department of Computer Science and Technology, Tsinghua University
Email: cuip@tsinghua.edu.cn
Tel: (+86)10-6279-0810