

Lei Li

RESEARCH INTERESTS

machine learning, data mining, natural language understanding and generation.

EDUCATION

Carnegie Mellon University, Pittsburgh, Pennsylvania USA

Ph.D., Computer Science Department, Sep. 2011

Dissertation: Fast algorithms for mining co-evolving time series. (recipient of ACM SIGKDD best dissertation award(runner-up))

M.S. in Science, May. 2009.

Advisor: Christos Faloutsos

Shanghai Jiao Tong University, China

B.S., Computer Science (ACM class), Sep, 2002 - Jun, 2006

Thesis: Semantic Search based on Probabilistic Description Logic Program.

Advisor: Yong Yu

EMPLOYMENT

ByteDance AI Lab

ByteDance

Research Scientist and Director

since 2016

machine writing, machine translation, (near-dup) video search, creative ads, ML for user growth

We develop a robot writer: Xiaomingbot, which integrates machine learning, NLP, data mining to automatically create human-quality news articles in multiple domains and languages. It is covered by many media reports.

Baidu USA

Sunnyvale, CA

Principal Research Scientist

Oct, 2014 - 2016

Institute of Deep Learning, Baidu Research

focus on deep learning and natural language understanding

University of California, Berkeley

Postdoctoral Researcher in EECS department

Oct, 2011 - Oct, 2014

working with Stuart Russell

Bayesian inference for open universe probability models. Compiler for probabilistic programs.

Carnegie Mellon University

Visiting Researcher in Machine Learning Department

Oct, 2012 - Aug, 2013

working with Eric Xing

IBM T.J. Watson Research Center, Hawthorne, New York USA

Intern at Healthcare Transformation group.

May, 2010 - Aug, 2010

Microsoft Research, Redmond, Washington USA

Intern at Networked Embedded Computing group.

Jun, 2009 - Aug, 2009

Google, Mountain View, California USA

Intern at Adspam group.

May, 2008 - Aug, 2008

APEX Data & Knowledge Management Lab, Shanghai, China

Research Assistant, on *semantic search* and *semantic web*

Jul, 2004 - Jun, 2006

Microsoft Research Asia, Beijing, China

Intern, on *semantic file system* project.

Aug, 2005 - Oct, 2005

Three patents granted: Patent No. US7,502,785, US 7,634,471 and US 7,624,130

PUBLICATIONS

REFEREED JOURNAL PAPERS

- [J1] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, **Lei Li**, and Christos Faloutsos. Non-linear dynamics of information diffusion in social networks. *ACM Transactions on the Web*, 11(1), February 2017.
- [J2] **Lei Li**, Bin Fu, and Christos Faloutsos. Efficient parallel learning of hidden markov chain models on smps. *IEICE Transactions on Information and Systems*, E93.D(6):1330–1342, 2010.
- [J3] **Lei Li**, B. Aditya Prakash, and Christos Faloutsos. Parsimonious linear fingerprinting for time series. *The Proceedings of the Very Large Data Bases Endowment (VLDB)*, 3:385–396, September 2010.
- [J4] Fan Guo, **Lei Li**, Christos Faloutsos, and Eric P. Xing. C-dem: a multi-modal query system for drosophila embryo databases. *The Proceedings of the Very Large Data Bases Endowment (VLDB)*, 1:1508–1511, August 2008.

REFEREED CONFERENCE AND WORKSHOP PAPERS

- [C1] Xunpeng Huang, Xianfeng Liang, Zhengyang Liu, **Lei Li**, and Yue Yu. SPAN: A stochastic projected approximate newton method. In *the 34th AAAI Conference on Artificial Intelligence (AAAI)*, February 2020.
- [C2] Xinlong Wang, Wei Yin, Tao Kong, Yuning Jiang, **Lei Li**, and Chunhua Shen. Task-aware monocular depth estimation for 3d object detection. In *the 34th AAAI Conference on Artificial Intelligence (AAAI)*, February 2020.
- [C3] Qingyang Wu, **Lei Li**, Hao Zhou, Ying Zeng, and Zhou Yu. Importance-aware learning for neural headline editing. In *the 34th AAAI Conference on Artificial Intelligence (AAAI)*, February 2020.
- [C4] Jiacheng Yang, Mingxuan Wang, Hao Zhou, Chengqi Zhao, Weinan Zhang, Yong Yu, and **Lei Li**. Towards making the most of bert in neural machine translation. In *the 34th AAAI Conference on Artificial Intelligence (AAAI)*, February 2020.
- [C5] Ning Miao, Hao Zhou, Chengqi Zhao, Wenxian Shi, and **Lei Li**. Kernelized bayesian softman for text generation. In *the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, December 2019.
- [C6] Zhichen Zhao, **Lei Li**, Bowen Zhang, Meng Wang, Yuning Jiang, Li Xu, Fengkun Wang, and Weiyang Ma. What you look matters: Offline evaluation of advertising creatives for cold start problem. In *the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, November 2019.
- [C7] Mingxuan Wang, Jun Xie, Zhixing Tan, Jinsong Su, Deyi Xiong, and **Lei Li**. Towards linear time neural machine translation with capsule networks. In *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, November 2019.
- [C8] Qing-Yuan Jiang, Yi He, Gen Li, Jian Lin, **Lei Li**, and Wu-Jun Li.. SVD: A large-scale short video dataset for near duplicate video retrieval.. In *International Conference on Computer Vision (ICCV)*, October 2019.
- [C9] Xin Wang, Jiawei Wu, Junkun Chen, **Lei Li**, Yuan-Fang Wang, and William Yang Wang. VATEX: A large-scale, high-quality multilingual dataset for video-and-language research. In *International Conference on Computer Vision (ICCV)*, October 2019. Oral Presentation.

- [C10] Yao Fu, Hao Zhou, Jiaze Chen, and **Lei Li**. Rethinking text attribute transfer: A lexical analysis. In *the 12th International Conference on Natural Language Generation (INLG)*, October 2019.
- [C11] Yunfei Lu, Linyun Yu, Peng Cui, Chengxi Zang, Renzhe Xu, Yihao Liu, **Lei Li**, and Wenwu Zhu. Uncovering the co-driven mechanism of social and content links in user churn phenomena. In *the 25th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, New York, NY, USA, August 2019. ACM.
- [C12] Zhaoyue Sun, Jiaze Chen, Hao Zhou, Deyu Zhou, **Lei Li**, and Mingmin Jiang. Graspnookey: Automatic chinese commentary generation for snooker videos. In *the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 6569–6571, August 2019. Demos.
- [C13] Rongxiang Wen, Hao Zhou, Shujian Huang, Yifan Xia, **Lei Li**, and Jiajun Chen. Correct-and-memorize: Learning to translate from interactive revisions. In *the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, August 2019. Oral Presentation.
- [C14] Yu Bao, Hao Zhou, Shujian Huang, **Lei Li**, Lili Mou, Olga Vechtomova, Xinyu Dai, and Jiajun Chen. Generating sentences from disentangled syntactic and semantic spaces. In *the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, July 2019.
- [C15] Lin Qiu, Yunxuan Xiao, Yanru Qu, Hao Zhou, **Lei Li**, Weinan Zhang, and Yong Yu. Dynamically fused graph network for multi-hop reasoning. In *the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, July 2019.
- [C16] Huangzhao Zhang, Ning Miao, Hao Zhou, and **Lei Li**. Generating fluent adversarial examples for natural languages. In *the 57th Annual Meeting of the Association for Computational Linguistics (ACL) - short papers*, July 2019.
- [C17] Hao Wu, Jiayuan Mao, Yufeng Zhang, Weiwei Sun, Yuning Jiang, **Lei Li**, and Wei-Ying Ma. Unified visual-semantic embeddings: Bridging vision and language with structured meaning representations. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. Oral Presentation.
- [C18] Ning Miao, Hao Zhou, Lili Mou, Rui Yan, and **Lei Li**. CGMH: constrained sentence generation by metropolis-hastings sampling. In *the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, January 2019.
- [C19] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Yitan Li, and **Lei Li**. BRITS: Bidirectional recurrent imputation for time series. In *the 32nd Conference on Neural Information Processing Systems (NeurIPS)*, December 2018.
- [C20] Haoyue Shi, Hao Zhou, Jiaze Chen, and **Lei Li**. On tree-based neural sentence modeling. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, October 2018.
- [C21] Gen Li, Shikun Xu, Xiang Liu, **Lei Li**, and Changhu Wang. Jersey number recognition with semi-supervised spatial transformer network. In *IEEE Conference on Computer Vision and Pattern Recognition workshops, Computer Vision in Sports*, pages 1864–1871, June 2018.
- [C22] Jiawei Wu, **Lei Li**, and William Yang Wang. Reinforced co-training. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 1252–1262, . Association for Computational Linguistics, June 2018.
- [C23] Yusuf B. Erol, Yi Wu, **Lei Li**, and Stuart Russell. A nearly-black-box online algorithm for joint parameter and state estimation in temporal models. In *the 31st AAAI Conference on Artificial Intelligence (AAAI)*, July 2017.
- [C24] Yi Wu, **Lei Li**, Stuart J. Russell, and Rastislav Bodik. Swift: Compiled inference for probabilistic programming languages. In *25th International Joint Conference on Artificial Intelligence (IJCAI)*, July 2016.
- [C25] Zihang Dai, **Lei Li**, and Wei Xu. Cfo: Conditional focused neural question answering with large-scale knowledge bases. In *ACL*, July 2016.

- [C26] Zefu Lu, **Lei Li**, and Wei Xu. Twisted recurrent network for named entity recognition. In *Bay Area Machine Learning Symposium*, 2015.
- [C27] Hieu Pham, Zihang Dai, and **Lei Li**. On optimization algorithms for recurrent networks with long short-term memory. In *Bay Area Machine Learning Symposium*, 2015.
- [C28] Yi Wu, **Lei Li**, and Stuart J. Russell. Bfit: From possible-world semantics to random-evaluation semantics in open universe. In *Neural Information Processing Systems, Probabilistic Programming workshop*, 2014.
- [C29] Simon Shaolei Du, Yilin Liu, Boyi Chen, and **Lei Li**. Maxios: Large scale nonnegative matrix factorization for collaborative filtering. In *Neural Information Processing Systems, workshop on Distributed Machine Learning and Matrix Computations*, 2014.
- [C30] Da-Cheng Juan, **Lei Li**, Huan-Kai Peng, Diana Marculescu, and Christos Faloutsos. Beyond poisson: Modeling inter-arrival times of requests in a datacenter. In *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2014.
- [C31] **Lei Li**, Bharath Ramsundar, and Stuart Russell. Dynamic scaled sampling for deterministic constraints. In *16th International Conference on Artificial Intelligence and Statistics*, 2013.
- [C32] Sharad Vikram, **Lei Li**, and Stuart Russell. Handwriting and gestures in the air, recognizing on the fly. In *ACM Conference on Human Factors in Computing Systems (CHI) Extended Abstracts*, 2013.
- [C33] Siyuan Liu, **Lei Li**, and Ramayya Krishnan. Hibernating process: Modelling mobile calls at multiple scales. In *IEEE International Conference on Data Mining*, 2013.
- [C34] Yusuf Erol, **Lei Li**, Bharath Ramsundar, and Stuart J. Russell. The extended parameter filter. In *Proceedings of the 30th International Conference on Machine learning*, 2013.
- [C35] Bin Fu, Jialiu Lin, **Lei Li**, Christos Faloutsos, Jason Hong, and Norman Sadeh. Why people hate your app - making sense of user feedback in a mobile app store. In *the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, New York, NY, USA, 2013. ACM.
- [C36] Mark Rogers, **Lei Li**, and Stuart J. Russell. Multilinear dynamical systems for tensor time series. In *the 27th Conference on Neural Information Processing Systems*, 2013.
- [C37] Yasuko Matsubara, **Lei Li**, Evangelos E. Papalexakis, David Lo, Yasushi Sakurai, and Christos Faloutsos. F-trail: Finding patterns in taxi trajectories. In *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 86–98, 2013.
- [C38] Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Hanghang Tong, Sugato Basu, Leman Akoglu, Danai Koutra, Christos Faloutsos, and **Lei Li**. Rolx: Structural role extraction and mining in large graphs. In *KDD '12: Proceeding of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2012. ACM.
- [C39] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, **Lei Li**, and Christos Faloutsos. Rise and fall patterns of information diffusion: Model and implications. In *KDD '12: Proceeding of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2012. ACM.
- [C40] Siyuan Liu, **Lei Li**, Christos Faloutsos, and Lionel Ni. Mobile phone graph evolution: Findings, model and interpretation. In *IEEE International Conference on Data Mining, workshop on Data Mining Technologies for Computational Collective Intelligence*, 2011.
- [C41] Keith Henderson, Brian Gallagher, **Lei Li**, Leman Akoglu, Tina Eliassi-Rad, Hanghang Tong, and Christos Faloutsos. It’s who you know: Graph mining using recursive structural features. In *KDD '11: Proceeding of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2011. ACM.
- [C42] **Lei Li**, Chieh-Jan Mike Liang, Jie Liu, Suman Nath, Andreas Terzis, and Christos Faloutsos. Thermo-cast: A cyber-physical forecasting model for data centers. In *KDD '11: Proceeding of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2011. ACM.

- [C43] **Lei Li** and B. Aditya Prakash. Time series clustering: Complex is simpler!. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [C44] Yasushi Sakurai, **Lei Li**, Yasuko Matsubara, and Christos Faloutsos. Windmine: Fast and effective mining of web-click sequences. In *SIAM International Conference on Data Mining*, 2011.
- [C45] **Lei Li**. Fast algorithms for time series mining. In *26th IEEE International Conference on Data Engineering, PHD Workshop*, pages 341–344, 2010.
- [C46] Keith Henderson, Tina Eliassi-Rad, Christos Faloutsos, Leman Akoglu, **Lei Li**, Koji Maruhashi, B. Aditya Prakash, and Hanghang Tong. Metric forensics: a multi-level approach for mining volatile graphs. In *KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 163–172, New York, NY, USA, 2010. ACM.
- [C47] **Lei Li**, James McCann, Nancy Pollard, and Christos Faloutsos. Bolero: a principled technique for including bone length constraints in motion capture occlusion filling. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '10, pages 179–188, Aire-la-Ville, Switzerland, Switzerland, 2010. Eurographics Association.
- [C48] **Lei Li**, James McCann, Nancy Pollard, and Christos Faloutsos. Dynammo: Mining and summarization of coevolving sequences with missing values. In *KDD '09: Proceeding of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2009. ACM.
- [C49] Fan Guo, **Lei Li**, and Christos Faloutsos. Tailoring click models to user goals. In *Proceedings of the 2009 workshop on Web Search Click Data*, WSCD '09, pages 88–92, New York, NY, USA, 2009. ACM.
- [C50] Wanhong Xu, Xi Zhou, and **Lei Li**. Inferring privacy information via social relations. In *IEEE 24th International Conference on Data Engineering workshops*, pages 525–530, 2008.
- [C51] **Lei Li**, Wenjie Fu, Fan Guo, Todd C. Mowry, and Christos Faloutsos. Cut-and-stitch: efficient parallel learning of linear dynamical systems on smps. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 471–479, New York, NY, USA, 2008. ACM.
- [C52] Yasushi Sakurai, Rosalynn Chong, **Lei Li**, and Christos Faloutsos. Efficient distribution mining and classification. In *SIAM International Conference on Data Mining*, pages 632–643, 2008.
- [C53] **Lei Li**, James McCann, Christos Faloutsos, and Nancy Pollard. Laziness is a virtue: Motion stitching using effort minimization. In *The 29th Annual Conference of the European Association for Computer Graphics, Short Paper Proceedings*, 2008.
- [C54] **Lei Li**, Qiaoling Liu, Yunfeng Tao, Lei Zhang, Jian Zhou, and Yong Yu. Providing an uncertainty reasoning service for semantic web application. In *Asia-Pacific Web Conference*, pages 628–639, 2006.

TECHNICAL REPORTS

- [T1] **Lei Li** and Stuart J. Russell. The blog language reference. Technical report UCB/EECS-2013-51, EECS Department, University of California, Berkeley, May 2013.
- [T2] **Lei Li**. *Fast algorithms for mining co-evolving time series*. PhD thesis, Carnegie Mellon University, 2011.

MEDIA COVERAGE

1. Disrupted Journalism, DisruptionHub. July 20, 2017. (mentioning our in-house developed Xiaomingbot) [link](#)
2. Robot gives competitive edge to tech application, China Daily, September 4, 2016. (similar coverage by 83 reports. about our news writing robot: Xiaomingbot, some examples see below)
3. Prolific Robot Journalist Covers 450 Olympic Stories. Newsweek. August 26, 2016. (coverage of our in-house developed writing robot Xiaomingbot) [link](#)
4. The Future of Writing? China's AI Reporter Published 450 Articles During Rio Olympics, Futurism, August 26, 2016. [link](#)

5. Robot Reporter Covered The Olympics For A Chinese News Outlet, Popular Science, August 25, 2016. (coverage of our in-house developed writing robot Xiaomingbot) [link](#)
6. China Sends AI Reporter to Cover the Rio Olympics – Blows Every Journalist Out of the Water, Nextshark, August 25, 2016. [link](#)
7. A Chinese news outlet used an incredibly efficient “robot reporter” to cover the Olympics. Quartz. August 24, 2016. [link](#)
8. China’s A.I. Olympics Reporter Could Replace Journalists, Inverse, August 24, 2016. [link](#)
9. Breaking down China’s most popular news app, Toutiao, AbacUSNews, Oct. 2, 2018. (mentioning Xiaomingbot) [link](#)
10. The Insanely Popular Chinese News App That You’ve Never Heard Of, MIT Technology Review, January 26, 2017. (Mentioning Xiaomingbot) [link](#)

PATENTS

1. Systems and methods for human inspired simple question answering, filed 2016.05, US20170109355A1, (pending)
2. Extracting semantic attributes, Patent No. US7,502,785, issued Mar 2009.
3. System and method for exploring a semantic file network, Patent No. US 7,624,130, issued Nov, 2009.
4. Adaptive grouping in a file network, Patent No. US 7,634,471, issued Dec 2009.

TALKS

1. Invited Talk: Deep Text Generation: Models and Applications, School of Data Science and Engineering, East China Normal University, Shanghai, China, November 8, 2019.
2. Industrial Invited Talk at CIKM 2019: Multi-lingual cross-modal machine translation at global scale, Beijing, China, November 7, 2019.
3. Tutorial at EMNLP 2019: Discreteness in Natural Language Processing, Hong Kong, November 4, 2019. (3 hr, with Lili Mou and Hao Zhou)
4. CCF-ADL107/NLPCC 2019 Tutorial: Deep Generative Models for Text Generation, Dunhuang, China, Oct 9, 2019. (3 hr, with Hao Zhou).
5. Invited Talk at IJCAI 2019 Industrial Days: Deep Text Generation: Models and Applications, Macau, China, August 2019.
6. Invited Talk: Machine Writing and Constrained Text Generation, CCF-AI workshop at Beijing Jiao Tong University, June 2019.
7. Invited Talk: Machine Writing and Constrained Sequence Generation, Arizona State University, Tempe, AZ, April 2019.
8. Invited Talk, Constrained Sentence Generation, School of Data Science, Fudan University, Shanghai, March 2019.
9. Invited Talk: Recent Progress in Natural Language Representation Learning and Generation, InfoQ AIcon, Beijing, December 21, 2018.
10. Invited Talk: Learning to generate better news, translation, and advertisement. CSDN AI NEXTCon, Beijing, November 13, 2018.
11. Invited Talk: AI for Content Creation and Interaction, Department of Computer Science, University of Illinois Urbana Champaign, November 2018.
12. Invited Talk: AI for Content Creation and Interaction, Viterbi School of Engineering, University of Southern California, October 2018.
13. Invited Talk: Incubating Innovative IT Talents - an academic-industry fused perspective. Computer Education Conference, CCF, Hefei, China, July 2018.
14. Invited Talk: AI for Content Creation and Discovery, Computer Science Department, University of California at Davis, California, April, 2018.

15. Invited Talk: Machine Writing and Language Generation, Computer Science Department, Fudan University, March, 2018.
16. Invited Talk: AI for Content Creation and Interaction, ShanghaiTech University, March, 2018
17. Invited Talk: AI for Content Creation and Interaction, Southeast University, March, 2018
18. Invited Talk: AI for Content Creation and Interaction, Nanjing University, March, 2018
19. Invited Talk: AI for Content Creation and Interaction, Nanjing Normal University, March, 2018
20. Invited Talk: AI-assisted Content Creation, CAAI MIFS (Machine Intelligence Frontier Symposium), Beijing, October, 2017.
21. Invited Talk: AI-Assisted Content Creation and Commenting, Institute of Computing at China Academy of Science, Beijing, September, 2017.
22. Invited Talk: AI-powered Information Creation, Distribution, and Interaction, AI Computing Conference, Beijing, September, 2017.
23. Invited Talk: AI Research at Toutiao, Data Science in China, KDD, Halifax, 2017.
24. CCF-ADL/NLPCC 2016 Tutorial: Deep Learning for Answering Questions, Kunming, China, December 2016.
25. Invited Talk: Learning to Understand and Generate Language, AIWorld Conference, Beijing, October, 2016.
26. Invited Talk: Status and Challenges towards general AI. China Conference on Artificial Intelligence. Beijing, August 26-27, 2016.
27. Invited Talk: AI for Content Creation and Sharing, CCF-GAIR (Global Conference on AI and Robotics), Shenzhen, August, 2016.
28. Invited Talk: Learning for Dialog and Question Answering, Shanghai Jiao Tong University, Shanghai, June, 2016.
29. Invited Talk: Large-scale Learning for Language Understanding, University of Science and Technology in China, Hefei, May, 2016.
30. Invited Talk: Learning for Language and Visual Understanding, CAAI-GAITC (Global AI Technology Conference), Beijing, April, 2016.
31. Invited Keynote: Deep Learning – Towards More Intelligent Machines, China Computer Federation Young Computer Scientists and Engineers Forum, April, 2015.
32. BLOG language and compiled inference. Computer Science department, Stanford University, and separately at Tsinghua University 2015.
33. BLOG language and efficient inference. Berkeley Programming Language Retreat, Santa Cruz, 2014
34. Scalable Probabilistic Inference for Complex Dynamical Models. Invited talk at Ohio State University, New York University, Emory University, Google (Mountain View), 2014.
35. Invited Fish Bowl Seminar: Fast Algorithms for Mining Co-evolving Time Series. , Texas A&M University, 2013.
36. DBLOG: an Open Universe Probabilistic Programming Language for Relational Modeling, The Second EITA Young Investigator Conference, Palo Alto, CA, Jul 2012.
37. Forecasting with Cyber-physical Interactions in Data Centers. CMU Parallel Data Lab. Sep, 2011.
38. Proactive Detection of Insider Threats with Graph Analysis and Learning (CMU representative), DARPA ADMAS program kickoff conference, Arlington, VA. Jun, 2011.
39. Fast Algorithms for Mining Co-evolving Time Series. Case Western Reserve University, PARC, Yahoo!, SIAT 2011.
40. Tutorial: Indexing and Mining Time Sequences. Washington DC, KDD 2010. (with Christos Faloutsos)
41. Parsimonious Linear Fingerprinting for Time Series. CMU machine learning lunch, 2010.
42. PLiF: Parsimonious Linear Fingerprinting for Time Series. Zhejiang University, China. Sep, 2010.
43. Fast Algorithms for Mining Co-evolving Time Series. SMU, NUS, HKUST, SJTU, Dec 2009.
44. Mining with Missing Values in Coevolving Time Series. MSR redmond, May 2009.
45. Efficient Parallel Learning of Linear Dynamical Systems on SMPs. CMU machine learning lunch, 2008.

AWARDS

1. 2019 Young Elite Award by China Computer Federation.
2. 2017 2nd-class WU Wenjun Award in AI for the project Internet information summarization and machine writing key technologies (including AI writer Xiaomingbot), the premier award by China Association of Artificial Intelligence.
3. 2017 China Computer Federation Distinguished Speaker.
4. 2012 SIGKDD Doctoral Dissertation Award (runner-up), the Association for Computing Machinery.

TEACHING EXPERIENCE

1. Guest Lecture, Deep Generative Models for Text Generation, graduate class in Machine Learning, Tsinghua University, November 26, 2019.
2. Guest Lecture, Deep Natural Language Generation, graduate class in Machine Learning, Tsinghua University, May 15, 2019.
3. Guest Lecture, Deep Natural Language Generation, graduate class in NLP, Peking University, May 8, 2019.
4. Guest Lecture, Personalized Recommendation Systems in Practice, undergrad Data Mining class, Tsinghua University. April 26, 2019
5. Guest Lecture, Deep Learning for Language Understanding and Generation, graduate Machine Learning class, Tsinghua University. November 13, 2018.
6. Guest Lecture, Deep Learning for Language Understanding and Generation, graduate Machine Learning class, Tsinghua University. April 25, 2018.
7. Guest Lecture, Deep Learning for Question Answering, graduate Machine Learning class, Tsinghua University. March 13, 2017.
8. Lecturer in Probabilistic Programming for Advanced Machine Learning Summer School, Portland, OR, USA, 2014. (total 12hrs lecture during two weeks)
9. Teaching Assistant for
 - 10701 Machine Learning, instructor Ziv Bar-Joseph, CMU, Spring 2009.
 - 15415 Database Applications, instructor Christos Faloutsos, CMU, Fall 2008.
 - Modern DBMS Implementation, SJTU, Fall 2005.
 - Operating System Project(nachos), SJTU, Fall 2005.
 - Modern Compiler Implementation, SJTU, Spring 2005.

STUDENTS ADVISING

1. Danlu Chen (student from Fudan, intern at Bytedance, Oct 2019 - , to join UCSD as Ph.D. student)
2. Jingjing Gong (Ph.D. student from Fudan, intern at Bytedance, July 2019 - now)
3. Zewei Sun (master student from NJU, intern at Bytedance, June 2019 - now)
4. Zhenqiao Song (master student from Fudan, intern at Bytedance, June 2019 - now)
5. Yuxuan Song (master student from SJTU, intern at Bytedance, June 2019 - now)
6. Rong Ye (master student from Fudan, intern at Bytedance, April 2019 - now)
7. Chi Sun (master student from Fudan, intern at Bytedance, April 2019 - now)
8. Lin Qiu (Ph.D. student from SJTU, intern at Bytedance, Jan 2019 - Dec 2019)
9. Yu Bao (Ph.D. student from NJU, intern at Bytedance, Jan 2019 - Nov 2019)
10. Qi Zeng (intern at Bytedance, summer 2019, joint w/ Hao Zhou, currently Ph.D. student at UIUC)
11. Yao Fu (intern at Bytedance Dec. 2017 - Jun. 2018, PKU undergrad, now master student at Columbia University, to be Ph.D. student at University of Edinburgh)
12. Shiyi Han (intern at Bytedance, now Ph.D. student at Brown University)
13. Freda(Haoyue) Shi (intern at Bytedance, joint w/ Hao Zhou, PKU undergrad, now Ph.D. student at TTIC)

14. Wenxian Shi (master student at PKU, intern at Bytedance April 2018 - June 2019, joint w/ Hao Zhou, currently SDE at Bytedance)
15. Ning Miao (master student at PKU, intern at Bytedance March 2018 - June 2019, joint w/ Hao Zhou, currently SDE at Bytedance)
16. Bohan Li (ZJU undergrad, intern at Bytedance 2017, now Ph.D. student at CMU LTI)
17. Ying Zeng (master student from PKU, intern at Bytedance Oct 2017 - June 2018, currently SDE at Bytedance)
18. Chengqi Zhao (master student from NJU, intern at Bytedance June 2017 - September 2017, currently SDE at Bytedance)
19. Gen Li (master student from CAS, intern at Bytedance April 2016 - December 2016, currently SDE at Bytedance)
20. Fei Jiang (Tsinghua, April 2015 - Oct 2015, now at JP Morgan Chase China)
21. Zefu Lu (Intern at Baidu, UIUC undergrad, May 2015 - Aug 2015)
22. Zihang Dai (Intern at Baidu, CMU, June 2015 - Aug 2015, now Ph.D. student at CMU)
23. Hieu Pham (intern at Baidu, Stanford master, June 2015 - Sep 2015, now Ph.D. student at CMU)
24. Di Zhou (U. Texas, May 2015 - Aug 2015)
25. Da Tang (UCB/Tsinghua undergrad, intern at UCB, July 2014 - Nov 2014, now Ph.D. student at Columbia U.)
26. Lea Jia (UCB undergrad, intern Feb 2014 - May 2014, Google)
27. Daniel Liu (UCB undergrad, intern Feb 2014 - Oct 2014)
28. Tejas Srinivasan, (UCB undergrad, intern Oct 2013 - May 2014)
29. Boyi Chen (UCB undergrad, intern Sep 2013 - Oct 2014, currently software engineer at LinkedIn)
30. Christopher Xie (UCB undergrad, intern Aug 2013 - May 2014 currently Ph.D. student at UW)
31. Yi Wu (Tsinghua undergrad, currently UCB Ph.D. student, intern at UCB, Aug 2013 -2014 (joint w/ Stuart Russell), summer intern 2016 at Bytedance)
32. Simon Du (UCB undergrad, intern Jul 2013 - Jan 2015, now Ph.D. student at CMU)
33. Zhiting Hu (CMU/Peking undergrad, intern Jul 2013 - Aug 2013, currently Ph.D. student at CMU)
34. Hao Fu (CMU/Tsinghua undergrad, Feb 2013 - May 2013)
35. Yiyao Jiang (CMU/Tsinghua undergrad, Feb 2013 - May 2013)
36. Bharath Ramsundar (UCB/Stanford undergrad, Jun 2012 - Dec 2013, currently Ph.D. student at Stanford)
37. Gerald Fong (UCB undergrad, Sep 2012 - May 2013, currently software engineer at Dropbox)
38. Sharad Vikram (UCB undergrad, Aug 2012 - May 2013, currently Ph.D. student at UCSD)
39. Aaron Wong (UCB master, Jul 2012 - June 2013, current software engineer at Google)
40. Cheng Lai Victor Huang (UCB undergrad, Sep 2012 - Jan 2013)
41. Huanhuan Wang (CMU master, joint w/ Siyuan Liu, sp 2012, currently at Google)
42. Shenghao Huang (CMU master, joint w/ Siyuan Liu, sp 2012, currently at LinkedIn)
43. Mark Rogers (UCB master, Jan 2012 - Dec 2013, joint w/ Stuart Russell)
44. Akihiro Matsukawa (UCB undergrad, Oct 2011 - Jan 2012, joint w/ Stuart Russell, currently software engineer at Twitter)

PROFESSIONAL SERVICE

SENIOR PC AND ORGANIZERS:

1. AAAI 2020 SPC
2. KDD Sponsorship Co-Chair 2020
3. EMNLP 2019 Area Chair, NLP Applications and Text Mining
4. CIKM 2019 SPC, Industrial Track
5. KDD Sponsorship Co-Chair 2019

6. AAAI 2019 SPC
7. CCAI (China Conference on Artificial Intelligence) 2019, Chair for Forum on Innovative Applications of AI
8. KDD Hands-on Tutorial Co-Chair 2018
9. IJCAI 2017 SPC
10. KDD Cup Co-Chair 2017
11. CCAI (China Conference on Artificial Intelligence) 2017, Co-Chair for Young Scientist Forum
12. CAAI-IEEE MIFS (Machine Intelligence Frontier Symposium) Co-Chair, 2016
13. CAAI Byte Tech AI Symposium, PC Co-Chair, Beijing, January 2019.
14. Organizer for 2018 Byte Cup International Machine Learning Competition. Task: News Title Generation. 1072 teams with 2175 participants from 35 countries. (Hosted by CAAI, ByteDance, & Tsinghua Knowledge Engineering Lab), [link](#)
15. Organizer for 2016 Byte Cup International Machine Learning Competition. Task: Expert Recommendation for Community Questions. 1000 teams with 1600 participants. (Hosted by CAAI, ByteDance, & Tsinghua Knowledge Engineering Lab), [link](#)

PROGRAM COMMITTEE:

1. KDD 2015, 2016, 2019.
2. ICDM 2015 Demos
3. ICML 2014
4. ECMLPKDD 2014, 2015
5. SDM 2013, 2014
6. IJCAI 2011, 2013, 2016, 2017, 2019
7. AAAI 2017, 2018, 2019
8. NeurIPS 2016
9. CIKM 2017
10. NAACL 2018
11. COLING 2016
12. EMNLP 2019
13. NLPCC 2019

JOURNAL REVIEWER: Transactions on Multimedia 2014, TVCJ 2014, Neural Computing 2013, TOSN 2013, 2014, TKDE 2012 - 2016, SADM 2012, Data Mining and Knowledge Discovery 2011, 2012, 2014, TOMCCAP 2009

PROPOSAL REVIEWER: Reviewer for US National Science Foundation proposal, 2010.