Lei Li

Research Interests

machine learning, 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 growth

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

Carnegie Mellon University, Pittsburgh, Pennsylvania USA

Research Assistant in Database group **Aug, 2006 - Oct, 2011**

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

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. The earlier version of the paper appeared in KDD'12. This version includes significant extension..

[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. This one is applying the idea from Cut-And-Stitch paper from linear dynamical system to hidden markov models. The extended version can be found in my thesis Chapter 6 and 7..

[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.

Conference Papers

[C1] 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, 2018.

[C2] 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)*, 2018.

[C3] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Yitan Li, and **Lei Li**. Brits: Bidirectional recurrent imputation for time series. In *Neural Information Processing Systems (NIPS)*, 2018.

[C4] 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, 2018.

[C5] 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 Thirty-First AAAI Conference on Artificial Intelligence*, 2017. The earlier version appeared in NIPS 2016 workshop on Advances in Approximate Bayesian Inference..

[C6] Zihang Dai, **Lei Li**, and Wei Xu. Cfo: Conditional focused neural question answering with large-scale knowledge bases. In *ACL*, 2016.

[C7] 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)*, 2016.

[C8] Zefu Lu, **Lei Li**, and Wei Xu. Twisted recurrent network for named entity recognition. In *Bay Area Machine Learning Symposium*, 2015.

[C9] 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.

[C10] 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.

[C11] 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.

[C12] 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.

[C13] 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. The full version appeared as Tech. Rep. UCB/EECS-2013-48..

[C14] 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 *KDD '13: Proceeding of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2013. ACM.

[C15] **Lei Li**, Bharath Ramsundar, and Stuart Russell. Dynamic scaled sampling for deterministic constraints. In *16th International Conference on Artificial Intelligence and Statistics*, 2013.

[C16] Siyuan Liu, **Lei Li**, and Ramayya Krishnan. Hibernating process: Modelling mobile calls at multiple scales. In *IEEE International Conference on Data Mining*, 2013.

[C17] 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.

[C18] Mark Rogers, **Lei Li**, and Stuart J. Russell. Multilinear dynamical systems for tensor time series. In *Advances in Neural Information Processing Systems 26*, 2013.

[C19] 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.

[C20] 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.

[C21] 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.

[C22] 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.

[C23] **Lei Li**, Chieh-Jan Mike Liang, Jie Liu, Suman Nath, Andreas Terzis, and Christos Faloutsos. Thermocast: 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.

[C24] **Lei Li** and B. Aditya Prakash. Time series clustering: Complex is simpler!. In *Proceedings of the 28th International Conference on Machine learning*, 2011. Please see for updated and additional experiments in Chap 5 of the thesis "Fast algorithms for mining co-evolving time series"..

[C25] 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.

[C26] 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.

[C27] 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.

[C28] **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.

[C29] **Lei Li**. Fast algorithms for time series mining. In *26th IEEE International Conference on Data Engineering, PHD Workshop*, pages 341–344, 2010.

[C30] 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.

[C31] **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.

[C32] **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.

[C33] **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.

[C34] 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.

[C35] 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.

[C36] **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.

Workshop & posters:

Mark Rogers, Lei Li, Stuart Russell. Dynamic Tensor Systems for Weather Prediction. Workshop on Understanding Climate Change. 2012 (Mark got student travel award)

Fast Algorithms for Mining Co-evolving Time Series. KDD 2011, Doctoral Event. San Diego, CA. 2011. (poster, recipient of travel grant)

Fast Algorithms for Mining Co-evolving Motion Capture Sequences. SDM 2010 Doctoral Forum. Columbus, OH. (poster, recipient of travel grant)

Lei Li. Fast Algorithms for Time Series Mining. ICDE 2010 Ph.D workshop. Long Beach, LA, 2010.

(recipient of travel fellowship)

DynoMMo: Mining with Missing Values in Coevolving Time Series. Parallel Data Lab retreat 2009 (poster).

Media Coverage

1. Robot Reporter Covered The Olympics For A Chinese News Outlet, Popular Science, August 25, 2016.
2. China Sends AI Reporter to Cover the Rio Olympics âĂŤ Blows Every Journalist Out of the Water, Nextshark, August 25, 2016.
3. Robot gives competitive edge to tech application, China Daily, September 4, 2016. (similar coverage by 83 reports.)
4. The Insanely Popular Chinese News App That YouâĂŹve Never Heard Of, MIT Technology Review, January 26, 2017.

Patents

1. Extracting semantic attributes, Patent No. US7,502,785, issued Mar 2009.
2. System and method for exploring a semantic file network, Patent No. US 7,624,130, issued Nov, 2009.
3. Adaptive grouping in a file network, Patent No. US 7,634,471, issued Dec 2009.

Talks

1. China Computer Federation Advanced Disciplines Lectures: Deep Learning for Answering Questions, December 2016.
2. Invited Talk: Status and Challenges towards general AI. CCAI. August, 2016.
3. Invited Keynote: Deep Learning – Towards More Intelligent Machines, China Computer Federation Young Computer Scientists and Engineers Forum, April, 2015.
4. BLOG language and compiled inference. Computer Science department, Stanford University, Tsinghua University 2015.
5. BLOG language and efficient inference. Berkeley Programming Language Retreat, Santa Cruz, 2014
6. Scalable Probabilistic Inference for Complex Dynamical Models. Invited talk at Ohio State University, New York University, Emory University, Google (Mountain View), 2014.
7. Invited Fish Bowl Seminar: Fast Algorithms for Mining Co-evolving Time Series. , Texas A&M University, 2013.
8. DBLOG: an Open Universe Probabilistic Programming Language for Relational Modeling, The Second EITA Young Investigator Conference, Palo Alto, CA, Jul 2012.
9. Forecasting with Cyber-physical Interactions in Data Centers. CMU Parallel Data Lab. Sep, 2011.
10. Proactive Detection of Insider Threats with Graph Analysis and Learning (CMU representative), DARPA ADMAS program kickoff conference, Arlington, VA. Jun, 2011.
11. Fast Algorithms for Mining Co-evolving Time Series. Case Western Reserve University, PARC, Yahoo!, SIAT 2011.
12. Parsimonious Linear Fingerprinting for Time Series. CMU machine learning lunch, 2010.
13. PLiF: Parsimonious Linear Fingerprinting for Time Series. Zhejiang University, China. Sep, 2010.
14. Fast Algorithms for Mining Co-evolving Time Series. SMU, NUS, HKUST, SJTU, Dec 2009.
15. Mining with Missing Values in Coevolving Time Series. MSR redmond, May 2009.
16. Efficient Parallel Learning of Linear Dynamical Systems on SMPs. CMU machine learning lunch, 2008.

Awards

1. 2012 SIGKDD Doctoral Dissertation Award (runner-up), the Association for Computing Machinery.
2. 1st place, China National Olympiad in Informatics, Jiangsu, 2001. (full score)
3. 1st class award, China National Physics Olympiad, Jiangsu, 2001.

Professional Service

Program Committee: KDD 2015, 2016. KDD Cup Co-Chair 2017. ICDM 2015 Demos. ICML 2014. ECMLPKDD 2014, 2015. SDM 2013, 2014. IJCAI 2011, 2013, 2016, 2017. AAAI 2017. ICDM 2011 workshop on collective intelligence. KDD workshop on Multimedia data mining(2011, 2012, 2013).

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.