

Lei Li

CONTACT INFORMATION

Sutardja Dai Hall 738
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720

E-mail: leili@cs.berkeley.edu
WWW: www.cs.cmu.edu/~leili

RESEARCH INTERESTS

machine learning and data mining in general:
communication/social network and societal data analysis
data intensive/cloud enable computing
time series analysis, with applications in motion capture, healthcare, data center monitoring, environmental monitoring
probabilistic programming language, reasoning and inference in open universe models

EDUCATION

Carnegie Mellon University, Pittsburgh, Pennsylvania USA

Ph.D., Computer Science Department, Sep. 2011
Dissertation: Fast algorithms for mining co-evolving time series.
M.S. in Science, May. 2009.
Advisor: Christos Faloutsos

Shanghai Jiao Tong University, China

B.S., Computer Science, Sep, 2002 - Jun, 2006
Thesis: Semantic Search based on Probabilistic Description Logic Program.
Advisor: Yong Yu

EXPERIENCE

University of California, Berkeley

Postdoctoral Researcher in EECS department
working with Stuart Russell

since Oct, 2011 - present

Carnegie Mellon University, Pittsburgh, Pennsylvania USA

Research Assistant in Database group
Time series learning with missing values, with application on motion capture (with Nancy Pollard), networking traffic modeling, and sensor stream monitoring.
Data mining in Fly embryo gene image database (with Eric Xing).

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.
on mining sensor data for saving energy in data centers.

Jun, 2009 - Aug, 2009

Google, Mountain View, California USA

Intern at Adspam group.
on anomaly detection in advertisement click spams.

May, 2008 - Aug, 2008

APEX Data & Knowledge Management Lab, Shanghai, China

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

Jul, 2004 - Jun, 2006

build a novel search model on semantic web, integrating information retrieval and logic reasoning techniques, to capture both formal semantics from knowledge bases and information from texts.

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] **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.
- [J2] **Lei Li**, B. Aditya Prakash, and Christos Faloutsos. Parsimonious linear fingerprinting for time series. *Proc. VLDB Endow.*, 3:385–396, September 2010.
- [J3] Fan Guo, **Lei Li**, Christos Faloutsos, and Eric P. Xing. C-dem: a multi-modal query system for drosophila embryo databases. *Proc. VLDB Endow.*, 1:1508–1511, August 2008.

CONFERENCE PAPERS

- [C1] 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.
- [C2] **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.
- [C3] **Lei Li** and B. Aditya Prakash. Time series clustering: Complex is simpler!. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.
- [C4] 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.
- [C5] Yasushi Sakurai, **Lei Li**, Yasuko Matsubara, and Christos Faloutsos. Windmine: Fast and effective mining of web-click sequences. In *SDM*, 2011.
- [C6] 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 *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.
- [C7] **Lei Li**. Fast algorithms for time series mining. In *ICDE 2010 PHD Workshop*, pages 341–344, 2010.
- [C8] **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.
- [C9] 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.
- [C10] **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.
- [C11] Yasushi Sakurai, Rosalynn Chong, **Lei Li**, and Christos Faloutsos. Efficient distribution mining and classification. In *SDM*, pages 632–643, 2008.

- [C12] **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.
- [C13] **Lei Li**, James McCann, Christos Faloutsos, and Nancy Pollard. Laziness is a virtue: Motion stitching using effort minimization. In *Short Papers Proceedings of EUROGRAPHICS*, 2008.
- [C14] 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.
- [C15] **Lei Li**, Qiaoling Liu, Yunfeng Tao, Lei Zhang 0007, Jian Zhou, and Yong Yu. Providing an uncertainty reasoning service for semantic web application. In *APWeb*, pages 628–639, 2006.

TECHNICAL REPORTS

- [T1] **Lei Li**. *Fast algorithms for mining co-evolving time series*. PhD thesis, Carnegie Mellon University, 2011. Available as technical report CMU-CS-11-127.

IN SUBMISSION:

Yasuko Matsubara, Lei Li, Evangelos Papalexakis, David Lo, Yasushi Sakurai, Christos Faloutsos, Mining real taxi trajectories, using fractals and tensors, 2012.

Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Leman Akoglu, Danai Koutra, Lei Li, Sugato Basu, Christos Faloutsos, "RoIX: Structural Role Extraction and Mining in Large Networks, 2012.

WORKSHOP & POSTERS:

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).

TALKS

Forecasting with Cyber-physical Interactions in Data Centers. CMU Parallel Data Lab. Sep, 2011.

Proactive Detection of Insider Threats with Graph Analysis and Learning (CMU representative), DARPA ADMAS program kickoff conference, Arlington, VA. Jun, 2011.

Fast Algorithms for Mining Co-evolving Time Series. Case Western Reserve University, PARC, Yahoo!, SIAT 2011.

Parsimonious Linear Fingerprinting for Time Series. CMU machine learning lunch, 2010.

PLiF: Parsimonious Linear Fingerprinting for Time Series. Zhejiang University, China. Sep, 2010. (*invited talk*)

Fast Algorithms for Mining Co-evolving Time Series. SMU, NUS, HKUST, SJTU, Dec 2009. (*invited talk*)

Mining with Missing Values in Coevolving Time Series. MSR redmond, May 2009.

Efficient Parallel Learning of Linear Dynamical Systems on SMPs. CMU machine learning lunch, 2008.

AWARDS

Student travel award, KDD 2011.

Student fellowship, ICML 2011. (\$1,000)

Travel grant for SDM, 2010.

Travel fellowship for ICDE, 2010. (\$ 1,000. sponsored by NSF IIS-0956600)

Three exceptional students, Shanghai Jiao Tong University, 2005. (top 1%)

Second place, China Undergraduate Mathematical Contest in Modeling, 2004

Tung's scholarship, 2004, 2005.

Exceptional Student award of Shanghai Jiao Tong University, 2003 (top 1%)

1st place, China National Olympiad in Informatics, Jiangsu, 2001. (full score)

1st class award, China National Physics Olympiad in Jiangsu, 2001.

2nd class award, China National High School Mathematics contest. 2000(Xiwangbei), 2001.

Huaying Scholarship, City of Changzhou, 2000, 2002.

PROFESSIONAL SERVICE

PROGRAM COMMITTEE: IJCAI 2011. ICDM 2011 workshop on collective intelligence.

CONFERENCE REVIEWER: KDD 2010, SIGMOD 2010, VLDB 2009, IJCAI 2009, PKDD 2008, WWW 2008, VLDB 2008, KDD 2007.

JOURNAL REVIEWER: Data Mining and Knowledge Discovery 2011 (three times), TOMCCAP 2009

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

ORGANIZER: CMU Database seminar series, 2009-2010.

VIDE-PRESIDENT: CMU Chinese Student & Scholar Association, 2008-2009.

CO-FOUNDER: Association of Astronomy. Shanghai Jiao Tong University. 2004.

OTHER PROJECTS

Approximate Parallel Learning for Sequential Graphical Models, Computer Architecture Fa07. Paper published in KDD.

Text Classifier with latent Dirichlet allocation(LDA), Machine Learning Fa06.
mainly responsible for the LDA part.

FatWorm DBMS with JDBC and RMI, Feb, 2005. (with Yaodong Zhang and Linji Yang)

A DBMS supports millions of records, SQL with subquery, transaction, multi-users at a time, logging, JDBC and remote access.

mainly responsible for the core architecture, B+ tree indexing, query evaluation and optimization, transaction, currency control, and JDBC implementation. Use a lot of Design Patterns.

Nachos OS with TCP, Jan, 2005.

An OS project from Berkeley, with a multi-threading kernel run on a simulated MIPS machine. I design the process scheduling, virtual memory and network transmission protocol.

Tiger Language Compiler, Jul, 2004.

COMPUTER SKILLS

Java(JSP/Servlet/RMI/JDBC/Struts), C/C++, Python, Matlab, R, OpenMP, SQL/Datalog compiler and optimizer, Tomcat+MySQL/PostgreSQL.