Zitao(Jerry) Liu

CONTACT INFORMATION

Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15213 USA

Voice: +1 (412) 614-0513
E-mail: ztliu@cs.pitt.edu
WWW: `http://www.cs.pitt.edu/~ztliu`

RESEARCH INTERESTS

Machine Learning, specialized in Time Series, Probabilistic Graphical Model and Optimization. Plentiful industrial experience in Forecasting, Text Mining, Recommendation and Ranking.

EDUCATION

University of Pittsburgh, Pittsburgh, Pennsylvania USA

Ph.D. Candidate, Computer Science Expected graduation date: May 2016

- Dissertation Topic: "Time Series Modeling of Irregularly Sampled Multivariate Clinical Data"
- Advisor: Milos Hauskrecht (milos@cs.pitt.edu) GPA: 3.88/4.0
- Google Scholar Profile: `http://scholar.google.com/citations?user=rRTzNmOAAAAJ`

Wuhan University, Wuhan, Hubei China

B.Eng., Software Engineering May, 2010 GPA: 3.71/4.0

HONORS & AWARDS

- Andrew Mellon Predoctoral Fellowship, University of Pittsburgh 2015
- People's Choice Graduate Poster Winner, University of Pittsburgh 2015
- Best Graduate Poster Runner-Up, University of Pittsburgh 2014
- SDM Student Travel Award 2013, 2014
- ICML Student Travel Award 2012
- Arts and Sciences Fellowship, University of Pittsburgh 2010
- Best Thesis Award of Bachelor's Degree, Ministry of Education of Hubei Province 2010
- National Scholarship, Ministry of Education of China 2007, 2009
- Citigroup Scholarship, Citigroup Services and Technology (China) Limited 2008

INDUSTRY EXPERIENCE

Alibaba Group, Seattle, WA

Research Intern, Institute of Data Science and Technology **06/2015 - 08/2015**

Designed and implemented a large scale user targeting and recommendation algorithm for Tmall (`http://www.tmall.com/`), which is the largest premier business-to-consumer online retail website in China. The algorithm is built on 10 billion highly sparse user behavior records (click, collect, cart, purchase information) involving 42 million users and 150 thousand shops. The algorithm includes a mix of collaborative filtering and pairwise learning to rank using logistic regression.

Yahoo! Lab, Sunnyvale, CA

Research Intern, Advertising Science Team **05/2014 - 08/2014**

Analyzed the web page traffic time series using seasonal-trend decomposition. Implemented an ensemble time series forecasting model which involves SVM Regression, Regression Tree, Gaussian Process, Gradient Boosting Tree, Local Regression, etc. Proposed an iterative algorithm to accurately estimate the missing values for hierarchical time series, which outperforms state-of-art missing value estimation techniques like matrix factorization, matrix completion, local regression, probabilistic PCA, etc.

ebay Research Lab, San Jose, CA

Research Intern, Data Science Team **06/2013 - 08/2013**
Performed a large scale query logs analysis for assessing personalization opportunities in eBay. Wrote Hadoop jobs to process 12 months (2012/08-2013/07) about 26 billion queries. Tried to answer questions like "What user information should we use in e-commerce websites to improve personalization?", "Should we do personalization on every query?", "Does recency effect will influence the personalized query prediction?"

Google, Mountain View, CA

Software Engineer Intern, Ads Review Team **06/2012 - 08/2012**
Built feature extractors for advertisements from Google AdWords, which involves more than 50 languages. Built and applied classifiers to detect illegal and malevolent advertisement texts.

Academic Experience

University of Pittsburgh, Pittsburgh, Pennsylvania USA

Missing Value Estimation for Hierarchical Time Series **11/2014 - 03/2015**
Studied the missing value estimation problem under hierarchical web traffic settings, where the user-visit traffics are organized in various hierarchical structures, such as geographical structure and website structure. Developed an iterative algorithm to accurately estimate the missing value in multivariate noisy web traffic time series with specific hierarchical structures by using subpsace and hierarchical consistency projections. (**ICDM 2015**)

Multivariate Time Series Analysis via Regularized Linear Dynamical Systems **05/2013 - 10/2014**
Proposed a regularized framework to (1) find the intrinsic latent state space dimensionality for multivariate time series, (2) prevent model overfitting given short multivariate time series and (3) support accurate forecasting. Applied various regularizers by introducing penalized priors and incorporate optimization techniques into the Maximum a Posteriori (MAP) learning framework. (**AAAI 2015**)

Multi-label Classification via Conditional Random Fields **09/2012 - 09/2013**
Achieved accurate multi-label classification in which data instances are associated with multiple, possibly high-dimensional, label vectors. Jointly learned the structure and parameters of Conditional Random Field. Maximized the pseudo likelihood of observed labels and used fast proximal gradient descend for learning the structure and limited memory BFGS for learning the parameters of the model. (**SDM 2014**)

Modeling Clinical Time Series Using Gaussian Process Sequences **06/2011 - 09/2012**
Developed a hierarchical framework to model irregularly sampled clinical time series and make accurate predictions. Combined advantages of linear dynamical systems and Gaussian processes. Modeled entire time series using a sequence of dependent Gaussian processes defined over time windows and captured the temporal dependence among Gaussian processes by using a linear dynamical system. (**SDM 2013**, **AIME 2013**)

Characterizing Machines and Workloads on a Google Cluster **09/2011 - 05/2012**
Studied a large-scale Google cluster usage trace dataset and characterized how the machines in the cluster are managed and the workloads submitted during a 29-day period behave. Focused on the frequency and pattern of machine maintenance events, job- and task-level workload behavior, and how the overall cluster resources are utilized. (**SRMPDS 2012**)

Subjectivity Feature Selection Package **01/2011 - 05/2011**
Implemented a Java-based package for sentiment classification (positive vs negative). Built a pipeline: part-of-speech tagging → negation handling → sentiment feature extraction (polarized unigrams/bigrams, adjective/adverb unigrams/bigrams/trigrams and transition features) → feature representation (TF-IDF and binary representations) → classification (Naive Bayes, kNN and SVM).

Wuhan University, Wuhan, Hubei China

Chinese News Classification Application **09/2009 - 05/2010**

Implemented a meta search engine prototype, which incorporates and merges the search results from Google, Baidu, Youdao, etc. Implemented an ensemble classifier for Chinese daily news using C#. Pipeline is built: Chinese word segmentation → feature extraction (mutual information, chi square, etc.) → feature representation (TF-IDF) → classification (Naive Bayes, kNN, etc.).

Journal Publications

Zitao Liu and Milos Hauskrecht. Clinical Time Series Prediction: Towards A Hierarchical Dynamical System Framework. *Journal of Artificial Intelligence in Medicine Special Issue*, 2014.

Conference Publications

Zitao Liu, Yan Yan, Jian Yang and and Milos Hauskrecht. Missing Value Estimation for Hierarchical Time Series: A Study of Hierarchical Web Traffic. *The IEEE International Conference on Data Mining(ICDM)*, 2015.

Zitao Liu and Milos Hauskrecht. A Regularized Linear Dynamical System Framework for Multivariate Time Series Analysis. *The Twenty-Ninth AAAI Conference on Artificial Intelligence(AAAI)*, 2015.

Mahdi Pakdaman, Iyad Batal, **Zitao Liu**, CharmGil Hong and Milos Hauskrecht. An Optimization-based Framework to Learn Conditional Random Fields for Multi-label Classification. *SIAM International Conference on Data Mining(SDM)*, 2014.

Zitao Liu and Milos Hauskrecht. Clinical Time Series Prediction with a Hierarchical Dynamical System. *14th Conference on Artificial Intelligence in Medicine(AIME)*, 2013.

Zitao Liu, Lei Wu and Milos Hauskrecht. Modeling Clinical Time Series Using Gaussian Process Sequences. *SIAM International Conference on Data Mining(SDM)*, 2013.

Zitao Liu, Wenchao Yu, Wei Chen, Shuran Wang and Fengyi Wu. Short Text Feature Selection and Classification for Micro Blog Mining. *International Conference on Computational Intelligence and Software Engineering(CiSE)*, 2010.

Zitao Liu, Wenchao Yu, Yalan Deng, Yongtao Wang and Zhiqi Bian. A Feature Selection Method for Document Clustering Based on Part-of-Speech and Word Co-Occurrence. *Proceedings of International Conference on Fuzzy Systems and Knowledge Discovery(FSKD)*, 2010.

Yang Shen, **Zitao Liu**, Cheng Luo and Ye Li. Research on Social Network Based on Meta-search Engine. *Proceedings of Web Information Systems and Applications Conference(WISA)*, 2009.

Yang Shen, **Zitao Liu**, Shaoji Luo, Huijuan Fu and Ye Li. Empirical Research on E-Government Based on Content Mining. *Proceedings of International Conference on Management of E-Commerce and E-Government(ICMeCG)*, 2009.

Workshop & Posters

Zitao Liu and Yan Yan. A Probabilistic Framework for Hierarchical Time Series Forecasting. *The 32nd International Conference on Machine Learning Workshop on Demand Forecasting (ICML-Workshop)*, 2015.

Zitao Liu, Gyanit Singh, Nish Parikh and Neel Sundaresan. A Large Scale Query Logs Analysis for Assessing Personalization Opportunities in E-commerce Sites. *ACM International Conference on Web Search and Data Mining Workshop on Log-Based Personalization(WSDM-Workshop)*, 2014.

Zitao Liu and Milos Hauskrecht. Sparse Linear Dynamical System with Its Application in Multivariate Clinical Time Series. *Neural Information Processing Systems Workshop on Machine Learning for Clinical Data Analysis and Healthcare(NIPS-Workshop)*, 2013.

Zitao Liu, Lei Wu and Milos Hauskrecht. State Space Gaussian Process Prediction. *29th International Conference on Machine Learning Workshop on Clinical Data Analysis(ICML-Workshop)*, 2012.

Zitao Liu and Sangyeun Cho. Characterizing Machines and Workloads on a Google Cluster, *Proceedings of the Eighth International Workshop on Scheduling and Resource Management for Parallel and Distributed Systems(SRMPDS)*, 2012.

Yang Shen, Huijuan Fu, **Zitao Liu**, Pengpeng Liu and Qingchuan Fu. Empirical Analysis on Chinese Academic Plagiarism. *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries(JCDL-Poster)*, 2009.

Academia Services

PC Member

- 11th International Conference on Machine Learning and Data Mining
- 7th Asian Conference on Intelligent Information and Database Systems
- 12th, 13th International Conference on Signal Processing and Multimedia Applications

Journal Reviewer

- ACM Transactions on Intelligent Systems and Technology
- IEEE Transactions on Neural Networks and Learning Systems
- Applied Clinical Informatics
- Artificial Intelligence in Medicine
- Computers and Electrical Engineering
- International Journal of Science, Technology and Society

External Reviewer: ICML2015, NIPS2015, SDM2015, AAAI2015, CIKM2014, ICDM2014, ICDM2013

Teaching Experience

University of Pittsburgh, Pittsburgh, Pennsylvania USA

Teaching Assistant

- CS2610 Interface Design and Evaluation (Fall 2012)
- CS1571 Introduction to Artificial Intelligence (Fall 2012)
- CS2750 Machine Learning (Spring 2012)
- CS0441 Discrete Structures for Computer Science (Fall 2011, Spring 2013)
- CS0401 Intermediate Programming Using Java (Spring 2011)

Wuhan University, Wuhan, Hubei China

Teaching Assistant

- Introduction to Information System (Spring 2010)
- Introduction to Computer System (Fall 2009)