

Jun Zhu

Project Scientist
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213

Email: junzhu@cs.cmu.edu
Mobile: +1 (412) 926-6298
Homepage:
<http://www.cs.cmu.edu/~junzhu/>

Research Interests

Statistical Machine Learning, Probabilistic Graphical Models, Bayesian Nonparametrics, Sparse Learning in High Dimensions with Structured Input/Output, Web Data Mining, Computer Vision, and Social Network Analysis.

Education

- Ph.D in Computer Science, Tsinghua University, Beijing, P.R China, 2009.
Advisor: Prof. Bo Zhang
- B.S. in Computer Science, Tsinghua University, Beijing, P.R China, 2005.
Graduated with Honor.

Honors and Awards

- China Computer Federation (CCF) Distinguished Thesis Award, 2009;
- Distinguished Thesis Award of Tsinghua University, 2009;
- Innovation Fellow, CMU, 2009;
- First-class Graduate Scholarship, 2008;
- CMU Short-Term Visiting Scholarship, 2008
- Distinguished Graduate Student at DCST department, 2007;
- Microsoft Fellowship, 2007;
- First-class Jiangzhen Graduate Scholarship, 2006 & 2007;
- Best Research Intern, Microsoft Research Asia, 2006;
- Excellent Undergraduate Student, Tsinghua University, 2005;
- Conference Travel Award, ICML05, ICML07, ICML08, KDD06 & KDD07, NIPS10;

Working and Visiting Experience

Associate Professor starting from 2011.9
Tsinghua University Beijing, China
Work in Computer Science Department.

Project Scientist 2011.7 - present
Carnegie Mellon University Pittsburgh, PA, USA
Work with Prof. Eric P. Xing.

Post-Doctoral Fellow 2009.9 - 2011.6
Carnegie Mellon University Pittsburgh, PA, USA

Project: learning predictive and interpretable latent representations.

Brief Description: Work on developing novel models and efficient learning methods to discover predictive and interpretable latent representations from noisy and high-dimensional data, which range from text documents, natural scene images, to relational network graphs. Our techniques demonstrate the successful integration of several promising yet conceptually very different thoughts. For example, our methods MaxEnDNet (See JMLR09), MedLDA (See ICML09) and Infinite

SVM (iSVM) (ICML11, to appear) provide a successful way to integrate the ideas from Bayesian nonparametrics (or Bayesian inference in general), large-margin learning, and kernel machines to solve challenging learning problems, e.g., setting the number of components and learning flexible nonlinear local classifiers in a mixture-of-experts model; and learning structured prediction models in the presence of latent variables.

We also develop structured sparse coding methods for discovering topical representations. Compared to existing probabilistic topic models, our non-probabilistic formulations can be more effective in controlling posterior sparsity of inferred representations and more efficient in learning and inference with a simple coordinate descent algorithm, especially when considering non-trivial side information, such as discrete categorical labels and rich contextual features.

Worked with Prof. Eric P. Xing.

Visiting Researcher

Stanford University

2010.2 - 2010.3 & 2010.10

Stanford, CA, USA

Project: learning to represent images in a low-dimensional semantic latent space.

Brief Description: Work on developing a novel sparse coding method to automatically learn a low-dimensional latent space, into which images are projected from a high-level and over-complete representation. We extensively test the applicability of the new representations in various application scenarios, including scene classification, image retrieval and annotation, etc.

Worked with Prof. Li Fei-Fei.

Visiting Researcher

Carnegie Mellon University

2008.9 - 2009.2

Pittsburgh, PA, USA

Project: (1) sparse learning in Markov networks and (2) max-margin supervised topic models.

Brief Description: Work on the (1) sparse learning in max-margin Markov networks (M^3N s), and (2) max-margin supervised topic models. More specifically, (1) due to the KKT conditions, the max-margin Markov networks are dual sparse in the sense that only a few lagrange multipliers are non-zero. We suggest to learn a primal sparse (i.e., only a few input features have non-zero weights) max-margin Markov networks by using the ℓ_1 -norm regularization and the resulting model is called ℓ_1 - M^3N . We conducted theoretical and computational analysis of the connections between the ℓ_1 - M^3N and our proposed laplace max-margin Markov networks (Lap M^3N) (See ICML08). We also developed an efficient and robust EM-style algorithm to learn an ℓ_1 - M^3N ; (2) Current supervised topic models (e.g., latent Dirichlet allocation) apply maximum likelihood estimation. In this project, we proposed to perform the max-margin learning during the topic discovery process. This inter-play between max-margin learning and latent topic discovery leads to discovering more discriminative latent topics and improving the prediction (e.g., classification and regression) accuracy. This work will appear in ICML09 (2 papers) and KDD09.

Worked with Prof. Eric P. Xing.

Research Assistant

Microsoft Research, Asia

Summer 2008

Beijing, China

Project: Entity Relationship Identification

Brief description: In this project, we aim to extract the relationships between different entities, e.g., the “wife” and “husband” relationship between people. To extract entity relationships, supervised methods rely on a large number of manually tagged training examples and can’t be applied to web scale applications. The existing bootstrapping system *Snowball* applies heuristic rules to select extraction patterns. In our approach (i.e., StatSnowball), we adopt the bootstrapping framework of Snowball but select the patterns by doing the ℓ_1 -norm regularized maximum likelihood estimation. We apply the general Markov logic networks (MLNs) as our underlying probabilistic models. We demonstrate the advantages of our models on web entity relationship extraction. See Renlifang (in Chinese) for an application of entity relationship search; or EntityCube for an English demo. The details are presented in the WWW09 paper.

Worked with Dr. Zaiqing Nie, Ji-Rong Wen, and Lei Zhang.

Visiting Researcher
Carnegie Mellon University

2007.9 - 2008.3
Pittsburgh, PA, USA

Project: Maximum Entropy Discrimination Markov Networks

Brief description: We generalized the basic *maximum entropy discrimination* (MED) principle (Jaakkola et al., 1999) to the much broader structured prediction problems, and the new learning framework is called *Maximum Entropy Discrimination Markov Networks* (MaxEnDNet or MEDN). MaxEnDNet performs Bayesian-style learning with a prior, and it subsumes the standard max-margin Markov networks when the prior is a standard normal. Under this general framework, we particularly studied the Laplace MaxEnDNet (LapMEDN) (i.e., the prior is a Laplace distribution), which is also called the Laplace max-margin Markov Networks. The LapMEDN has a posterior shrinkage effect and thus is more suitable for learning Markov networks in sparse high dimensional spaces. We also generalized the MaxEnDNet to the *Partially Observed Maximum Entropy Discrimination Markov Networks* (PoMEN) to incorporate structured hidden variables. PoMEN can effectively learn a max-margin model from partially labeled data sets, for example, only the leaf nodes of a hierarchical model are labeled. See the ICML08, NIPS08 and JMLR09 papers for details.
Worked with Prof. Eric P. Xing.

Research Assistant
Microsoft Research Asia

2004.11 - 2007.7
Beijing, P.R China

Project: Statistical Web Data Extraction

Brief description: We aim to extract object information from web pages, such as people, products, and conferences. Traditional extractors heavily rely on the design templates of web pages. Thus, they are limited in the generalizability, and for different domains we have to develop different sets of extractors. In order to handle the huge number of web pages, we proposed to build statistical extraction models. Our models are probabilistic graphical models, particularly the Markov random fields. Under this statistical extraction framework, we particularly studied how to leverage the structural information of web pages to improve the extraction accuracy. Specifically, we developed the 2D conditional random fields (2D CRFs) for incorporating the 2D layout information of web pages, and developed the hierarchical conditional random fields (HCRFs) for incorporating the multi-level dependencies. We applied our models to the product information extraction and empirical results demonstrate their advantages. We have developed a Windows Live Product Search Engine to help users to find interested products for sale on the web. See the ICML05, KDD06, and KDD07 papers for more details.

Worked with Dr. Zaiqing Nie, Ji-Rong Wen, and Wei-Ying Ma.

Publications

1. **Jun Zhu**, Ning Chen, and Eric P. Xing. Infinite Latent SVM for Classification and Multi-task Learning. *To Appear in NIPS*, 2011.
2. **Jun Zhu** and Eric P. Xing. Sparse Topical Coding. *To Appear in UAI*, 2011.
3. **Jun Zhu**, Ni Lao, Ning Chen and Eric P. Xing. Conditioned Topical Coding: an Efficient Topic Model with Rich Features. *To Appear in SIGKDD* (full oral), 2011.
4. **Jun Zhu**, Ning Chen and Eric P. Xing. Infinite SVM: a Dirichlet Process Mixture of Large-margin Kernel Machines. *To Appear in ICML*, 2011.
5. **Jun Zhu**, Li-Jia Li, Li Fei-Fei and Eric P. Xing. Large Margin Learning of Upstream Scene Understanding Models. *In Proc. of NIPS*, 2010.
6. Ning Chen, **Jun Zhu** and Eric P. Xing. Predictive Subspace Learning for Multi-view Data: a Large Margin Approach. *In Proc. of NIPS*, 2010.
7. Seunghak Lee, **Jun Zhu** and Eric P. Xing. Adaptive Multi-Task Lasso: with Application to eQTL Detection. *In Proc. of NIPS*, 2010.
8. Ni Lao, **Jun Zhu**, Liu Liu, Yandong Liu and William Cohen. Efficient Relational Learning with Hidden Variable Detection. *In Proc. of NIPS*, 2010.

9. **Jun Zhu** and Eric P. Xing. Conditional Topic Random Fields. *In Proc. of ICML*, 2010.
10. **Jun Zhu**, Ni Lao, and Eric P. Xing. Grafting-Light: Fast, Incremental Feature Selection and Structure Learning of Markov Random Fields. *In Proc. of SIGKDD* (full oral), 2010.
11. **PhD Thesis**: Maximum Entropy Discrimination Markov Networks: Theory and Applications. Tsinghua University, Beijing China, July, 2009. (Awarded as **China Computer Federation (CCF) Distinguished Thesis; Distinguished Thesis of Tsinghua University**)
12. **Jun Zhu** and Eric P. Xing. Maximum Entropy Discrimination Markov Networks. *Journal of Machine Learning Research*, 10(Nov):2531-2569, 2009. (This is a full extension of the technical report and the Laplace maximum margin Markov networks).
13. **Jun Zhu** and Eric P. Xing. On the Primal and Dual Sparsity in Markov Networks. *In Proc. of ICML*, 2009.
14. **Jun Zhu**, Amr Ahmed, and Eric P. Xing. MedLDA: Maximum Margin Supervised Topic Models for Regression and Classification. *In Proc. of ICML*, 2009.
15. **Jun Zhu**, Eric P. Xing, and Bo Zhang. Primal Sparse Max-Margin Markov Networks. *In Proc. of SIGKDD* (full oral), 2009.
16. Xiaolin Shi, **Jun Zhu**, Rui Cai, and Lei Zhang. User Grouping Behavior in Online Forums. *In Proc. of SIGKDD* (short oral), 2009.
17. **Jun Zhu**, Zaiqing Nie, Xiaojiang Liu, Bo Zhang, and Ji-Rong Wen. StatSnowball: a Statistical Approach to Extracting Entity Relationships. *In Proc. of WWW* (full oral), 2009. (*The developed technique has been used in the Microsoft EntityCube project: <http://research.microsoft.com/enus/projects/entitycube/>.*)
18. Jiangming Yang, Rui Cai, Yida Wang, **Jun Zhu**, Lei Zhang, and Wei-Ying Ma. Incorporating Site-Level Knowledge to Extract Structured Data from Web Forums. *In Proc. of WWW* (full oral), 2009.
19. **Jun Zhu**, Zaiqing Nie, and Bo Zhang. Statistical Web Object Extraction. Invited Book Chapter in Encyclopedia of Data Warehousing and Mining, Second Edition, 2009.
20. **Jun Zhu**, Zaiqing Nie, Bo Zhang, and Ji-Rong Wen. Dynamic Hierarchical Markov Random Fields for Integrated Web Data Extraction. *Journal of Machine Learning Research*, 9(Jul):1583-1614, 2008.
21. **Jun Zhu**, Eric P. Xing, and Bo Zhang. Maximum Entropy Discrimination Markov Networks. *Technical Report, CMU-ML-08-104*, Carnegie-Mellon University, 2008.
22. **Jun Zhu**, Eric P. Xing, and Bo Zhang. Partially Observed Maximum Entropy Discrimination Markov Networks. *In Proc. of NIPS*, 2008.
23. **Jun Zhu**, Eric P. Xing, and Bo Zhang. Laplace Maximum Margin Markov Networks. *In Proc. of ICML*, 2008.
24. **Jun Zhu**, Zaiqing Nie, Bo Zhang, and Ji-Rong Wen. Dynamic Hierarchical Markov Random Fields and their Application to Web Data Extraction. *In Proc. of ICML*, 2007.
25. **Jun Zhu**, Zaiqing Nie, Ji-Rong Wen, Bo Zhang, and Hsiao-Wuen Hon. Webpage Understanding: an Integrated Approach. *In Proc. of SIGKDD* (full oral), 2007.
26. **Jun Zhu**, Zaiqing Nie, Ji-Rong Wen, Bo Zhang, and Wei-Ying Ma. Simultaneous Record Detection and Attribute Labeling in Web Data Extraction. *In Proc. of SIGKDD* (full oral), 2006. (*The developed technique has been used in the Windows Live Product Search project: <http://products.live.com/>.*)
27. **Jun Zhu**, Zaiqing Nie, Ji-Rong Wen, Bo Zhang, and Wei-Ying Ma. 2D Conditional Random Fields for Web Information Extraction. *In Proc. of ICML*, 2005.

Recent Talks

Invited Talks

- *Predictive Latent Subspace Learning: a Large Margin Approach*. CMU Machine Learning Lunch, 2010.
- *Maximum Entropy Discrimination Markov Networks: Theory and Applications*. Thesis work presented at Microsoft Research Asia, 2009; HP Lab, Beijing, 2009; NEC Laboratories America, Cupertino Campus, 2010; Stanford University, 2010.
- *MedLDA: Maximum Margin Supervised Topic Models for Regression and Classification*. Tsinghua NLP Seminar, 2009.
- *Partially Observed Maximum Entropy Discrimination Markov Networks*. CMU Machine Learning Lunch, 2008.
- *Laplace Maximum Margin Markov Networks*. Microsoft Research Asia, 2008.
- *Statistical Web Object Extraction*. IBM Research at Hawthorne, NY, 2008;
- *Statistical Webpage Modeling and Related Learning Problems*. Tsinghua Graduate Research Saloon, 2008.
- *An Introduction to Graphical Models*. Microsoft Research Asia, 2006.
- *An Introduction to Conditional Random Fields*. SIGMM Multimedia Seminar, Tsinghua, 2006.

Conference Talks

- *Infinite SVM: Dirichlet Process Mixture of Large-margin Kernel Machines*. ICML, 2011.
- *Grafting-Light: Fast, Incremental Feature Selection and Structure Learning of Markov Random Fields*. SIGKDD, 2010.
- *Conditional Topic Random Fields*. ICML, 2010.
- *MMH: Maximum Margin Supervised Harmoniums*. ICML workshop on topic models: Structure, Applications, Evaluation, and Extensions, 2010.
- *Primal Sparse Max-Margin Markov Networks*. SIGKDD, 2009.
- *MedLDA: Maximum Margin Supervised Topic Models for Regression and Classification*. ICML, 2009.
- *On Primal and Dual Sparsity in Markov Networks*. ICML, 2009.
- *Laplace Maximum Margin Markov Networks*. ICML 2008.
- *Webpage Understanding: an Integrated Approach*. SIGKDD 2007.
- *Simultaneous Record Detection and Attribute Labeling in Web Data Extraction*. SIGKDD 2006.

Professional Activities

Conference PC Member

- Advances in Neural Information Processing Systems (NIPS), 2011
- International Conference on Machine Learning (ICML), 2011
- International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2011
- International Conference on Data Engineering (ICDE), 2011
- International Conference on Empirical Methods in Natural Language Processing (EMNLP), 2011
- International Conference on Machine Learning (ICML), 2010
- Advances in Neural Information Processing Systems (NIPS), 2010
- International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2010
- Annual Meeting of the Association for Computational Linguistics (ACL), 2010
- AAAI Conference on Artificial Intelligence (AAAI), 2010

- International Conference on Empirical Methods in Natural Language Processing (EMNLP), 2010
- PhD Forum of Department of Computer Science and Technology. 2008, 2009.

Journal Reviewer

- Machine Learning Journal (MLJ), from 2008
- IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), from 2011
- Journal of Artificial Intelligence (JAIR), from 2009
- IEEE Trans. on Knowledge Discovery from Data (TKDD), from 2011
- Computational Statistics and Data Analysis (CSDA), from 2011
- International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), from 2011
- Information Sciences Journal, from 2010
- Information Systems Journal, from 2010
- Neural Computing Journal, from 2009
- Journal of Computer Science and Technology (JCST), from 2009
- Information Processing & Management Journal, from 2009

Other Activities

- Leader of SIGMM (Special Interest Group on Multi-media) Seminar, Tsinghua University, from 2006.9 to 2007.9.
- Student Volunteer for ICML 2005, 2008 and 2009; SIGKDD 2006, 2007 and 2009; NIPS 2008.

Patents

- Two-dimensional conditional random fields for web extraction. United States Patent 7529761. Inventors: Ji-Rong Wen; Wei-Ying Ma; Zaiqing Nie; Jun Zhu.
- Hierarchical conditional random fields for web extraction. United States Patent 7720830. Inventors: Ji-Rong Wen; Wei-Ying Ma; Zaiqing Nie; Jun Zhu.
- Web-Scale Entity Relationship Extraction. Pending. Inventors: Zaiqing Nie; Ji-Rong Wen; Jun Zhu

References

Available upon request.