

OBJECTIVE

A full-time research and development position in machine learning, data mining, natural language processing, and visual analytics.

EDUCATION

Purdue University, West Lafayette, IN

Ph.D. in Electrical and Computer Engineering, May 2009 (expected)

Advisor: Prof. Guy Lebanon; GPA: 3.85/4

Zhejiang University, Hangzhou, P.R.China

M.S. in Computer Science, Mar 2004

B.S. in Computer Science, Jun 2001

SKILLS

Machine Learning: information theory, probabilistic graphical models, convex optimization, models and algorithms for structured data

Statistics: constrained statistical inference, stochastic process, statistical models for ranked data

Natural Language Processing: document sentiment prediction, part-of-speech tagging, sequential modeling of text documents

Programming: C/C++, Matlab, Java, OpenGL

RESEARCH EXPERIENCE

Research Assistant, Purdue University, West Lafayette, IN Aug. 2004 – Present

- Developed a systematic way to obtain domain-dependent priors through probability elicitation and to incorporate them through parameter space regularization. Introduced isotonic conditional random fields (isotonic CRFs) in this general framework. Applied the isotonic CRFs to sentiment prediction, and demonstrated its promising usage in modeling the local sentiment flow, analyzing the author's writing style and summarizing the document content.
- Introduced a non-parametric estimator based on kernel smoothing for estimating the population distribution of partially ranked data. Derived an efficient computational method for some popular partial ranking types, making the estimator practical for real-world data sets, for example web-search data, movie review datasets, and multi-label text document datasets.
- Studied a sequential document representation called locally weighted bag of words (lowbow), which captures medium and long range sequential dependencies within a text document. Developed the practical applicability of lowbow to various text processing tasks including document classification, segmentation, summarization, and visualization.
- Explored statistical machine translation and diffusion kernels for unsupervised metric learning for text documents.
- Examined feature extraction from CT/MRI scans for effective volume rendering.

Research Assistant, Zhejiang University, Hangzhou, China Sep 2000 – Mar 2004

- Addressed the problem of part-based 3D model retrieval by using the distribution of shape index over the 3D model surface as the representation, and the earth mover's distance as the metric.
- Introduced a new shape descriptor for 3D polygon models.
- Developed an algorithm based on hidden Markov models for audio segmentation and classification.

INDUSTRIAL EXPERIENCE

Intern, Siemens Medical Solutions, Malvern, PA May 2008 – Aug 2008

- Explored contextual distribution, local smoothing and constrained random walk for estimating word similarity. Applied the similarity model for classifying medical documents, and for query expansion in searching medical document databases.

Intern, Microsoft Research Asia, Beijing, China

Jul 2001 – Sep 2001

- Studied the self-similarity property for a class of audio clips and utilized the property for audio synthesis and audio restoration.

TEACHING EXPERIENCE

Teaching Assistant for Data Structure (Fall 2005, Spring 2006)

Teaching Assistant for Probabilistic Methods for Electrical and Computer Engineering (Spring 2006)

HONORS AND AWARDS

Google Anita Borg Memorial Scholarship Finalist, 2008

Attendee of Google Workshop for Women Engineers, 2008

Student Travel Grant, ICML 2006

Rockwell Scholarship, Rockwell Automation, 2002

Graduated with University Distinction, Zhejiang University, 2001

Zhejiang University Scholarship, 1998-2000

Freshman Scholarship, 1997

SELECTED PUBLICATIONS

Journal Articles

Y. Mao and G. Lebanon, “Generalized Isotonic Conditional Random Fields,” submitted to *Machine Learning Journal*.

G. Lebanon and **Y. Mao**, “Non-Parametric Modeling of Partially Ranked Data,” *Journal of Machine Learning Research*, to appear.

G. Lebanon, **Y. Mao** and J. Dillon, “The Locally Weighted Bag of Words Framework for Document Representation,” *Journal of Machine Learning Research*, 8:2405–2441, 2007.

Y. Mao, J. Dillon and G. Lebanon, “Sequential Document Visualization,” *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1208–1215, 2007.

Y. Mao, H. Pan, F. Wu and Y. Zhuang, “Depth Weighted Normal Map for 3D Model Retrieval,” *Journal of Computer-Aided Design and Computer Graphics*, 17(2), 2005. (in Chinese)

Y. Zhuang, **Y. Mao**, F. Wu and Y. Pan, “Hidden Markov Model based Broadcast News Segmentation and Classification,” *Journal of Computer Research and Development*, 39(9), 2002. (in Chinese)

Conference Papers

G. Lebanon and **Y. Mao**, “Non-Parametric Modeling of Partially Ranked Data,” In *Advances in Neural Information Processing Systems*, 2008. (*oral*)

J. Dillon, **Y. Mao**, G. Lebanon and J. Zhang, “Statistical Translation, Heat Kernels, and Expected Distances,” In *The 23rd Conference on Uncertainty in Artificial Intelligence*, 2007. (*oral, nominated for best student paper award*)

Y. Mao and G. Lebanon, “Isotonic Conditional Random Fields and Local Sentiment Flow,” In *Advances in Neural Information Processing Systems*, 2007.

Y. Zhuang, **Y. Mao**, F. Wu and Y. Pan, “3D Model and Motion Retrieval: the Extended Dimensions for Web-Based Learning,” In *Lecture Notes in Computer Science* 2783, Springer Verlag, 2003 (also in *International Journal of Computer Processing of Oriental Languages*, 17(1), 2004).

L. Lu, **Y. Mao**, W. Liu and H. Zhang, “Audio Restoration by Constrained Audio Texture Synthesis,” In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003.