

JOSHUA V. DILLON

US Citizen

1305 Klaus Building, 266 Ferst Drive, Atlanta, GA 30332-0765
jvdillon@gatech.edu / +1 (765) 532-4643 / www.almostsure.com

OBJECTIVE

A full-time research or development position (Sept. '11) in machine learning, data mining, web search, on-line advertising, social networks, natural language processing, information retrieval or statistical data analysis.

EDUCATION

Georgia Institute of Technology, Atlanta, Georgia Jan. 2009 – May 2011

Ph.D., Computational Science & Engineering

Thesis: “Stochastic m-Estimators for Controlling Accuracy-Cost Tradeoffs in Machine Learning”

Advisor: Prof. Guy Lebanon

GPA: 4.00/4.00

Purdue University, West Lafayette, Indiana Aug. 2005 – Dec. 2008

M.S., Electrical & Computer Engineering

GPA: 3.69/4.00

Michigan Technological University, Houghton, Michigan Aug. 2001 – Apr. 2005

B.S., Computer Engineering & Electrical Engineering

Honors: Summa cum laude; Double major

GPA: 3.90/4.00

RESEARCH EXPERIENCE

Research Assistant, Georgia Institute of Technology, Atlanta, Georgia Jan. 2009 – Present

- Quantified the asymptotic accuracy of generative semi-supervised learning based on an extension of stochastic composite likelihood.
- Developed the stochastic m-estimator framework for controlling tradeoffs in machine learning such as computational cost, labeling cost, robustness. Examined large data statistical properties.
- Examined the effect of pseudo-periodicity in kernel density estimation and local likelihood methods and justified alternative techniques through analytical and empirical study.

Research Intern, Microsoft Research, Redmond, Washington Summer 2009

- Developed a flexible optimization framework for constraining probabilistic models with imprecise domain knowledge. Applied this framework for finding robust pseudo-relevance feedback models for information retrieval.

Research Assistant, Purdue University, W. Lafayette, Indiana Jan. 2006 – Dec. 2008

- Proposed a family of point estimators that resolve the computation-accuracy tradeoff present in maximum likelihood. Proved their consistency and provided formulas for their asymptotic variance and computational complexity. Demonstrated their usefulness for several graphical models including CRFs and Boltzmann machines.
- Developed the locally weighted bag of words framework for representing sequential text. Applied framework to several text analysis tasks, i.e., classification, segmentation, summarization, and visualization.
- Investigated statistical machine translation and diffusion kernels for unsupervised metric learning for text documents.

Summer Scholar, Lawrence Livermore Nat. Lab., Walnut Creek, California Summer 2006

- Investigated unsupervised learning techniques for statistical process control. Applied these techniques to the Joint Genome Institute’s DNA sequencing process to identify combinations of reagents, machines, and operators that lead to under-performing modes of operation.

Extreme Blue Intern, IBM, Austin, Texas Summer 2005

- Developed unsupervised classification techniques which exploit harmonically related features. Applied this work for automatic error detection in the Linux kernel. Feature engineering included, source/binary alignment, resolving dynamic control flow, and system profiling with call-stack resolution.

ENGINEERING EXPERIENCE

Intern, ThermoAnalytics Inc., Calumet, Michigan Spring 2005

- Solely designed and implemented a QScript-to-C translator optimized for numerical computing applications. Efforts included lexical analysis, context free grammar specification, developing an abstract syntax tree representation with corresponding auto-typing symbol table, and implementing the semantic (code-emitting) routines.

Intern, IBM, Rochester, Minnesota Summer 2004

- Implemented VHDL logic designs for the floating-point core of the Cell processor. Conducted timing analysis, synthesis, and testing of over 20 logic macros. Significantly improved team turnaround time by automating several report generating tasks (Perl) and implementing a tailored layout prototyping tool (Java).

Intern, IBM, Rochester, Minnesota Summer 2003

- Designed, implemented, and packaged an SAP R/3 cluster management plug-in for iSeries Navigator (Java). Design goals included extensibility, graphical ease-of-use, and an aggressive release cycle to meet clients' demands. Efforts also involved the coordination of domestic and German colleagues. End product was delivered in a fully packaged form, ahead of schedule.

Intern, Michigan Department of Transportation, Cass City, Michigan Summer 2002

- Sole on-site inspector responsible for verifying contractors' adherence to design specifications. Responsible for chemical and physical quality control, logging payable items, and updating project plans.

TEACHING EXPERIENCE

Lecturer, Purdue University, W. Lafayette, Indiana Spring, Fall 2008

- Two semester instructor for an ECE undergraduate course which acquaints students with scripted language software engineering tools, i.e., Python and korn shell. Responsibilities included curriculum design and delivering weekly lectures to 60+ students. One of only two graduate student lecturers.

Teaching Assistant, Purdue University, W. Lafayette, Indiana Fall 2007

- Coordinating TA for "Microprocessor System Design and Interfacing," an undergraduate course which introduces microprocessor system design, assembly programming, and digital/analog interfaces. Held lab office hours and managed five undergraduate TAs.

FELLOWSHIPS

- Marshall Sherfield Postdoctoral Fellowship, Marshall Aid Commemoration Commission, 2011-13.
- DHS Fellowship in Data Analysis and Visual Analytics, Dept. of Homeland Security, 2010-12.
- Ross Graduate Fellowship, Purdue University, 2005-06.
- Board of Control-Full Tuition, Michigan Technological University, 2001-05.

DISTINCTIONS

- US delegate, 57th Lindau Meeting of Nobel Laureates and Students, Germany, 2007.
- Summa cum laude, Dept. of Electrical and Computer Engineering, Michigan Tech. Univ., 2005.
- Award of Excellence, Mathematics Dept., Michigan Tech. Univ., 2002.

Other Distinctions

- Eta Kappa Nu ECE Honor Society, Beta Chapter. Purdue University, 2006-08.
- Eta Kappa Nu ECE Honor Society, Beta Gamma Chapter. Michigan Tech. Univ., 2005.
- Phi Kappa Phi Honor Society. Michigan Technological University, 2004.
- Tau Beta Pi Engineering Honor Society, Michigan Beta Chapter. Michigan Tech. Univ., 2004.
- Sheldon G. Hayes Foundation Scholar. Michigan Tech. Univ., 2003.
- Dr. C. M. Carson Memorial Scholar. Michigan Tech. Univ., 2001.
- Michigan Merit Award. State of Michigan, 2001.
- Valedictorian, Cass City High School. Cass City, Michigan, 2001.

SKILLS

Machine Learning: graphical models, convex optimization, info. theory, (semi-/un-) supervised learning
Statistics: point estimators, m-estimators, large sample theory, stochastic processes
Natural Language Processing: IR, classification, sequence labeling (POS, NER), sequential text models
Programming: Matlab, Perl, C++, C, Java, Python, Bash

SELECTED PUBLICATIONS

Journal Articles

- **J. Dillon** and G. Lebanon. Stochastic composite likelihood. *Journal of Machine Learning Research* (JMLR), 11(Sep):2597-2633, 2010.
- G. Lebanon, Y. Mao, and **J. Dillon**. The locally weighted bag of words framework for documents. *Journal of Machine Learning Research* (JMLR), 8(Oct):2405-2441, 2007.
- Y. Mao, **J. Dillon**, and G. Lebanon. Sequential document visualization. *IEEE Transactions on Visualization and Computer Graphics* (INFOVIS), 13(6):1208-1215, 2007.

Conference Papers

- **J. Dillon**, K. Balasubramanian, and G. Lebanon. Asymptotic analysis of generative semi-supervised learning. In *Proc. of the International Conference on Machine Learning* (ICML), 2010.
- **J. Dillon** and K. Collins-Thompson. A unified optimization framework for finding reliable pseudo-relevance feedback models. In *Proc. of the International Conference on Information and Knowledge Management* (CIKM), 2010.
- **J. Dillon** and G. Lebanon. Statistical and computational tradeoffs in stochastic composite likelihood. In *Proc. of the International Conference on Artificial Intelligence and Statistics* (AISTATS), 2009.
- **J. Dillon**, Y. Mao, G. Lebanon, and J. Zhang. Statistical translation, heat kernels, and expected distances. In *Uncertainty in Artificial Intelligence* (UAI), 2007.

Workshop Papers

- K. Collins-Thompson and **J. Dillon**. Controlling the search for expanded query representations by constrained optimization in latent variable space. *SIGIR Workshop on Query Representation and Understanding*, 2010.
- **J. Dillon**, Y. Mao, G. Lebanon, and J. Zhang. Statistical translation, heat kernels, and expected distances. In *Proc. of NIPS Workshop on Learning to Compare Examples*, 2006.

Software

- **J. Dillon**. Sharedmatrix. *Matlab Central*, August 2010.