Joseph E. Gonzalez

|  |  |
| --- | --- |
| Email: jegonzal@cs.cmu.edu  Phone: 626-394-2853 | <http://cs.cmu.edu/~jegonzal> |

My research addresses the challenges of designing and building large-scale machine learning algorithms and systems. In particular, my thesis work focuses on large-scale structured machine learning using probabilistic graphical models (Markov Random Fields) that are capable of reasoning about billions of related random variables. The resulting algorithms and systems have achieved state-of-the-art performance in tasks ranging from predicting ad preferences in social networks to solving complex protein modeling tasks. As part of my thesis work I created GraphLab (<http://graphlab.org>), a framework that dramatically simplifies the design and implementation of high-performance large-scale machine learning systems.

|  |  |  |
| --- | --- | --- |
| **Education** | | |
| **Carnegie Mellon University** | | Currently a 6th year PhD candidate in the Machine Learning Department advised by Carlos Guestrin. I will complete my PhD in August 2012. |
| **California Institute of Technology** | | B.S. in Computer Science with Honors, GPA 3.8 |
| **Publications** | | |
| * Amr Ahmed, Mohamed Aly, Joseph Gonzalez, Shravan Narayanamurthy, and Alex Smola. “*Scalable inference in latent variable models.”* WSDM, 2012. * Joseph Gonzalez, Yucheng Low, Arthur Gretton, and Carlos Guestrin. “*Parallel gibbs sampling: From colored fields to thin junction trees.”* AISTATS, 2011. * Joseph Gonzalez, Yucheng Low, and Carlos Guestrin. Scalable Machine Learning, Chapter *“Parallel Inference on Large Factor Graphs.”* Cambridge U. Press, 2010. * Y. Low, J. Gonzalez, A. Kyrola, D. Bickson C. Guestrin, J. Hellerstein. *“GraphLab: A New Parallel Framework for Machine Learning.”* UAI. 2010. * J. Gonzalez, Y. Low, C. Guestrin, and D. OHallaron. *“Distributed Parallel Inference on Large Factor Graphs.”* UAI. 2009 * J. Gonzalez, Y. Low, and C. Guestrin. *“Residual Splash for Optimally Parallelizing Belief Propagation.”* AISTATS. 2009 * R. Chamberlain, J. Gonzalez, G. Gutt, E Tailor*. “New Line of Sight Algorithm Renders Superlative TINs Superfluous”* JPL Document D-32587, *Export-Controlled, U.S. Gov’t Only.* | | |
| **Awards and Fellowships** | | |
| * **AT&T Labs Fellowship** **(2007):** Graduate research stipend for 3 years. * **NSF Graduate Research Fellowship (2007):** Graduate research stipend for 3 years. * **NASA Space Act Award (2005):** Awarded for a sizeable contribution to space exploration. * **NASA Inventions and Contributions Board Award (2005):** Awarded for the development of an innovative new technology that has made a sizeable contribution to space exploration. | | |
| **Work Experience** | | |
| **Work &**  **Research** | **Yahoo! Research (2011):** Extended the GraphLab abstraction to enable large-scale machine learning on natural graphs derived from social media and web-content. (Alex Smola: [smola@yahoo-inc.com](mailto:smola@yahoo-inc.com))  **AT&T Labs Research (2007):** Developed models for statistically assessing DSL quality from limited noisy data.(Steven Phillips: [phillips@research.att.com](mailto:phillips@research.att.com))  **ADAPT Automated AdWords Auction Agent (Spring 2006):** Developed and implemented models for assessing word value in the Google AdWords market. (Alex Bäcker: [alex@caltech.edu](mailto:alex@caltech.edu))  **Microsoft Developer Internship (2005):** Worked with MSN Search team developing techniques to use behavioral information to identity search spam. (Greg Hullender: [greghull@windows.microsoft.com](mailto:greghull@windows.microsoft.com))  **Undirected Search Algorithms (2004):** Developed a new query-less search technology that uses prior reading interests to identify novel documents. (Alex Bäcker: [alex@caltech.edu](mailto:alex@caltech.edu))  **Efficient Line-of-Sight Evaluation (2003):** Developed a new algorithm for efficiently evaluating line-of-sight on digital elevation maps at JPL. (Robert Chamberlain: [rgc@jpl.nasa.gov](mailto:rgc@jpl.nasa.gov)) | |
| **Skills** | **Advanced C++ & Systems Experience:** All of my current high-performance code is written in C++ using for linux with Boost, Lapack, Google Performance Tools, MPICH2, and Pthreads.  **Statistical Techniques:** Graphical models, statistical inference, nonparametric methods | |
| **Open Source**  **Software** | **GraphLab (C++):** A sophisticated API for building parallel and distributed  Machine Learning systems on top of multicore and cloud architectures. <http://graphlab.org>  **Parallel SplashBP & SplashGibbs (C++):** This library implements a collection of parallel statistical inference algorithms. <http://select.cs.cmu.edu/code/> | |