

## Research Summary

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 we created GraphLab, a framework that dramatically simplifies the design and implementation of high-performance large-scale machine learning systems.

For more information visit <http://cs.cmu.edu/~jegonzal>.

## Education

**PhD Candidate:** Machine Learning Department at Carnegie Mellon University.

**Advisor:** Carlos Guestrin

**Thesis Title:** “*Parallel Learning and Inference in Probabilistic Graphical Models.*”

**Anticipated Defense Date:** August 2012

**Masters:** Machine Learning Department at Carnegie Mellon University. [2010]

**Undergraduate:** B.S. in Computer Science from Caltech. [2006]

## Publications

- [1] Amr Ahmed, Mohamed Aly, Joseph Gonzalez, Shravan Narayanamurthy, and Alex Smola. Scalable inference in latent variable models. In *Conference on Web Search and Data Mining (WSDM)*, February 2012.
- [2] Joseph Gonzalez, Yucheng Low, Arthur Gretton, and Carlos Guestrin. Parallel gibbs sampling: From colored fields to thin junction trees. In *Artificial Intelligence and Statistics (AISTATS)*, May 2011.
- [3] Joseph Gonzalez, Yucheng Low, and Carlos Guestrin. *Scalable Machine Learning*, chapter Parallel Inference on Large Factor Graphs. Cambridge U. Press, 2010.
- [4] Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin, and J. M. Hellerstein. Graphlab: A new parallel framework for machine learning. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, July 2010.
- [5] J. Gonzalez, Y. Low, C. Guestrin, and D. O’Hallaron. Distributed parallel inference on large factor graphs. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, July 2009.

- [6] J. Gonzalez, Y. Low, and C. Guestrin. Residual splash for optimally parallelizing belief propagation. In *Artificial Intelligence and Statistics (AISTATS)*, April 2009.
- [7] R. Chamberlain, J. Gonzalez, G. Gutt, and E. Taylor. New line of sight algorithm renders superlative tins superfluous. Technical Report D-32587, JPL, 2005.
- [8] J. Gonzalez. Leaping over chasms. *CURJ*, 4, April 2004.

## Awards

- **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 contribution to space exploration.
- **Upper Class Merit Award (Twice) (2004,2005):** full tuition for academic excellence, research, and faculty recognition. Presidential Award (2002-2006): I was awarded full tuition for research and academic achievements.

## Scientific Community Involvement

- **[2011] NIPS Workshop Organizer** I organized and lead the workshop entitled “Big Learning: Algorithms, Systems, and Tools for Learning at Scale” For more information visit the workshop website <http://biglearn.org>
- **[2010] ICML Reviewer**
- **[2010] DARPA Future Ideas Symposium** Invited speaker at the DARPA future ideas symposium.
- **[2009] NIPS Workshop Organizer** I organized and lead the workshop entitled “Large-Scale Machine Learning: Parallelism and Massive Datasets.” For more information visit the workshop website <http://www.select.cs.cmu.edu/meetings/biglearn09>
- **[2009] JMLR Reviewer**
- **[2008 - 2009] Helped Lead a DARPA Interdisciplinary Sciences and Technology Study (ISAT) Group** to investigate the future of parallel machine learning from an interdisciplinary perspective I also participated in the final Woodshole annual ISAT meeting to prepare a proposal for the DARPA director.
- **[2007] JMLR Reviewer**
- **[2007] IPSN Reviewer**

## Teaching

- **[2009] Teaching Assitant** for the CMU Masters Machine Learning class. I designed problem sets and exams, gave recitation lectures, and mentored student projects. This was an ambitious class. (*Geoff Gordon: guestrin@cs.cmu.edu*)

- **[2007] Teaching Assistant** for the CMU Graduate Machine Learning class. I designed problem sets and exams, gave recitation lectures, and mentored student projects. (*Carlos Guestrin: guestrin@cs.cmu.edu*)
- **[2004,2005] Head Teaching Assistant** for the Caltech Introductory Computer Science Course (CS2): Redesigned introductory computer science. (*Al Barr: barr@cs.caltech.edu*)

## Industry Experience

- **Yahoo! Research (2011):** Developed the next generation of 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*)
- **AT&T Labs Research (2007):** Developed models for statistically assessing DSL quality from limited noisy data. (*Steven Phillips: phillips@research.att.com*)
- **Intern at ADAPT (2006)** Worked on an automated AdWords auction agent. I developed and implemented models for assessing word value. (*Alex Becker: alex@caltech.edu*)
- **Microsoft Developer Internship (2005):** Worked with MSN Search team developing techniques to use behavioral information to identify search spam. (*Greg Hullender: greghull@windows.microsoft.com*)
- **Caltech Research Fellowship (2004):** Developed a new query-less search technology that uses prior reading interests to identify novel documents. (*Alex Becker: alex@caltech.edu*)
- **NASA Jet Propulsion Labs Fellowship (2003):** Developed a new algorithm for efficiently evaluating line-of-sight on digital elevation maps at JPL. (*Robert Chamberlain: rgc@jpl.nasa.gov*)

## Publicly Released Software

- **GraphLab (C++)** A sophisticated API for building parallel and distributed machine learning algorithms on top of multicore and cloud architectures. GraphLab generalizes the MapReduce abstraction to support iterative asynchronous computation on graph structured dependent data. <http://graphlab.org>
- **Distributed SplashBP (C++)** This library implements the SplashBP algorithm for factor graph inference in the distributed setting using MPI. [http://www.select.cs.cmu.edu/code/mpi\\_splash.tar.gz](http://www.select.cs.cmu.edu/code/mpi_splash.tar.gz)
- **Shared Memory SplashBP (C++)** This library implements the SplashBP algorithm for Markov random fields inference. [http://www.select.cs.cmu.edu/code/parallelmrf\\_src.tar.gz](http://www.select.cs.cmu.edu/code/parallelmrf_src.tar.gz)