

GraphLab (All related to BigData)

Aapo Kyrola, Guy Blelloch, and Carlos Guestrin (2012). "GraphChi: Large-Scale Graph computation on Just a PC." Proceedings of the 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI '12).

Joseph E. Gonzalez, Yucheng Low, Haijie Gu, Danny Bickson, and Carlos Guestrin (2012). "PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs." Proceedings of the 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI '12).

Yucheng Low, Joseph Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin and Joseph M. Hellerstein (2012). "Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud." PVLDB.

Joseph K. Bradley, Aapo Kyrola, Danny Bickson, and Carlos Guestrin (2011). "Parallel Coordinate Descent for L1-Regularized Loss Minimization." International Conference on Machine Learning (ICML 2011).

Joseph Gonzalez, Yucheng Low, Arthur Gretton, and Carlos Guestrin (2011). "Parallel Gibbs Sampling: From Colored Fields to Thin Junction Trees." In Artificial Intelligence and Statistics (AISTATS).

Yucheng Low, Joseph Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin, and Joseph M. Hellerstein (2010). "GraphLab: A New Parallel Framework for Machine Learning." Conference on Uncertainty in Artificial Intelligence (UAI).

Joseph Gonzalez, Yucheng Low, Carlos Guestrin, and David O'Hallaron (2009). "Distributed Parallel Inference on Large Factor Graphs."

Conference on Uncertainty in Artificial Intelligence (UAI).

Joseph Gonzalez, Yucheng Low, and Carlos Guestrin (2009). "Residual Splash for Optimally Parallelizing Belief Propagation." In Artificial Intelligence and Statistics (AISTATS).

Graph Model Theory & Algorithm (More theoretical)

Joseph K. Bradley and Carlos Guestrin (2012). "Sample Complexity of Composite Likelihood." In Artificial Intelligence and Statistics (AISTATS). "We present the first PAC bounds for learning parameters of Conditional Random Fields with general structures over discrete and real-valued variables."

Le Song, Arthur Gretton, Danny Bickson, Yucheng Low, and Carlos Guestrin (2011). "Kernel Belief Propagation." In Artificial Intelligence and Statistics (AISTATS).

Anton Chechetka and Carlos Guestrin (2010). "Evidence-Specific Structures for Rich Tractable CRFs." In Advances in Neural Information Processing Systems (NIPS).
"We present a simple and effective approach to learning tractable conditional random fields with structure that depends on the evidence. Our approach retains the advantages of tractable discriminative models, namely efficient exact inference and arbitrarily accurate parameter learning in polynomial time. At the same time, our algorithm does not suffer a large expressive power penalty inherent to fixed tractable structures."

Joseph K. Bradley and Carlos Guestrin (2010). "Learning Tree Conditional Random Fields."

International Conference on Machine Learning (ICML 2010).

Le Song, Arthur Gretton, and Carlos Guestrin (2010). "Nonparametric Tree Graphical Models via Kernel Embeddings." In Artificial Intelligence and Statistics (AISTATS).

Anton Chechetka and Carlos Guestrin (2010). "Focused Belief Propagation for Query-Specific Inference." In Artificial Intelligence and Statistics (AISTATS). Best Student Paper Award.

Danny Bickson and Carlos Guestrin (2010). "Inference with Multivariate Heavy-Tails in Linear Models." Neural Information Processing Systems (NIPS).
"In this work, we propose a novel simple linear graphical model for independent latent random variables, called linear characteristic model (LCM), defined in the characteristic function domain. we show for the first time, how to compute both exact and approximate inference in such a linear multivariate graphical model."

Dafna Shahaf, Anton Chechetka, and Carlos Guestrin (2009). "Learning Thin Junction Trees via Graph Cuts." In Artificial Intelligence and Statistics (AISTATS).

Andreas Krause and Carlos Guestrin (2009). "Optimal Value of Information in Graphical Models." Journal of Artificial Intelligence Research (JAIR), 35, 557-591.

NLP + Knowledge Discovery (related to Big Data)

Dafna Shahaf, Carlos Guestrin, and Eric Horvitz (2012). "Metro Maps of Science." ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).

"In this paper, we propose metrics of influence, coverage, and connectivity for scientific literature. We use these metrics to create structured summaries of information, which we call metro maps."

Dafna Shahaf, Carlos Guestrin, and Eric Horvitz (2012). "Trains of Thought: Generating Information Maps." International World Wide Web Conference (WWW).

Khalid El-Arini and Carlos Guestrin (2011). "Beyond Keyword Search: Discovering Relevant Scientific Literature." ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).

Dafna Shahaf and Carlos Guestrin (2011). "Connecting Two (or Less) Dots: Discovering Structure in News Articles." ACM Transactions on Knowledge Discovery from Data.

Dafna Shahaf and Carlos Guestrin (2010). "Connecting the Dots Between News Articles." ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). Best Paper Award.

Khalid El-Arini, Gaurav Veda, Dafna Shahaf, and Carlos Guestrin (2009). "Turning Down the Noise in the Blogosphere." ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).

Recommendation System (deal with relatively small datasets)

Yisong Yue, Sue Ann Hong, and Carlos Guestrin (2012). "Hierarchical Exploration for Accelerating Contextual Bandits." International Conference on Machine Learning (ICML).

Yisong Yue and Carlos Guestrin (2011). "Linear Submodular Bandits and their Application to Diversified Retrieval." In Advances in Neural Information Processing Systems (NIPS).

Ranking Data Modeling (specific datasets)

Jonathan Huang, Ashish Kapoor, and Carlos Guestrin (2012). "Riffled Independence for Efficient Inference with Partial Ranking." Journal of Artificial Intelligence, 44, 491-532.

In this paper, we show that the recently proposed riffled independence assumptions cleanly and efficiently address each of the above challenges. In particular, we establish a tight mathematical connection between the concepts of riffled independence and of partial rankings.

Jonathan Huang and Carlos Guestrin (2012). "Uncovering the riffled independence structure of ranked data." Electronic Journal of Statistics, 6, 199-230.

Jonathan Huang, Ashish Kapoor, and Carlos Guestrin (2011). "Efficient Probabilistic Inference with Partial Ranking Queries." The 27th Conference on Uncertainty in Artificial Intelligence (UAI 2011).

Jonathan Huang and Carlos Guestrin (2010). "Learning Hierarchical Riffle Independent Groupings from Rankings." International Conference on Machine Learning (ICML 2010).

Jonathan Huang and Carlos Guestrin (2009). "Riffled Independence for Ranked Data." In Advances in Neural Information Processing Systems (NIPS).

Jonathan Huang, Carlos Guestrin, and Leonidas Guibas (2009). "Fourier Theoretic Probabilistic Inference over Permutations." Journal of Machine Learning Research (JMLR), 10, 997-1070.

Jonathan Huang, Carlos Guestrin, Xiaoye Jiang, and Leonidas Guibas (2009). "Exploiting Probabilistic Independence for Permutations." In Artificial Intelligence and Statistics (AISTATS).