
Romer E. Rosales Ph.D.

813 Williamsburg Blvd.
Downingtown, PA 19335

romer@csail.mit.edu
phone: (857) 222-7397

<http://www.csail.mit.edu/~romer>

Technical Expertise

Applied machine learning and data mining, with special focus on the analysis and modeling of mid to large-scale, real-world data and information sources as a means to understand and utilize this information in more valuable ways (e.g.; new computer tools, applications, business models, services). Development of algorithms and mathematical models primarily using probabilistic inference and convex optimization. Application domains include search/identification of concepts from free text, information extraction, health record analysis, medical imaging, gene sequence analysis, and computer vision. Recent areas of work include:

1. Information retrieval/extraction from free text: concept search rather than keyword search
2. Learning from crowds: learning from many non-expert annotators (or information sources)
3. Active learning/sensing: reducing data collection cost and improving annotator efficiency
4. Knowledge transfer: machine learning when training and real test distributions are different, in particular when unlabeled data is available
5. Combination of expert domain knowledge with learning: how to leverage available domain knowledge without requiring more data labeling
6. Approximate inference in graphical models: how to utilize large probability models even though basic (statistical inference) computations are intractable
7. ℓ_1 regularization for learning sparse models: how to learn parsimonious (simpler) classification/regression models efficiently through convex optimization
8. Estimating the structure of graphical models: automated probabilistic model building when little is known about the variable dependencies, with preference for simple structures

Academic Background

Research Associate Massachusetts Institute of Technology.	Feb 2004
Computer Science and Artificial Intelligence Laboratory (CSAIL). Cambridge, MA USA.	Oct 2004
Postdoctoral Research Fellow University of Toronto.	Jan 2002
Probabilistic and Statistical Inference Group. Electrical Engineering. Toronto, ON Canada.	Jan 2004
Ph.D. Computer Science Boston University.	Jan 2000
Computer Science Department. Boston, MA USA.	Dec 2002
M.A. Computer Science Boston University.	Sep 1996
Computer Science Department. Boston, MA USA.	Dec 1999
B.Eng. Informatics Engineering Universidad Centro Occidental <i>Lisandro Alvarado</i> .	Sep 1989
Faculty of Sciences. Venezuela. [ranked first in class - 1/123]	Jul 1995

Work Experience

Product/Project Manager	2009-Curr.
Senior Staff Scientist	2007-Curr.
Staff Scientist	2004-2007
IKM CAD and Knowledge Solutions. Siemens Healthcare. Malvern, PA	
Development of machine learning (ML) models and algorithms. Propose R&D directions to support the group's business plans and identify opportunities enabled by new/current technologies. For example, I proposed, help design, and currently manage a software project that combines ML, web services, and a specially tailored web GUI to ease product development (internal) and to enhance customer product functionality (external); this project is in the area of ML applied to search/identification of medical concepts in natural language text. Project lead, team size: 4-6 people. Coordinate	

interaction between text processing ML research and software platform. Basic business development responsibilities. Developed ML applications for analysis of medical text, images, and structured data from health records in general.

Other: Serve as principal coordinator of group's patent portfolio (responsible for 70-75 patent decisions/year). Coordinate recruiting for various technical positions. Organize Summer and International Internship programs for software engineering/machine learning students (undergraduate and graduate levels). Prepare grant applications.

Postdoctoral Associate

2004

**Computer Science and Artificial Intelligence Laboratory.
Massachusetts Institute of Technology. Cambridge, MA**

Designed and implemented novel approaches for approximate probabilistic inference in graphical models (probability models), called *Focused Inference* [21]. The main motivation was to solve probabilistic inference problems for applications with a large number of variables that interact in ways for which exact methods are not tractable.

Postdoctoral Fellow

2002-2003

Probabilistic and Statistical Inference Group. University of Toronto. Toronto, ON

Designed a novel multi-purpose method for general statistical image processing covering problems such as de-noising, super-resolution, non-photorealistic rendering, edge detection, texture transfer/generation. Developed various approaches for machine learning for pattern recognition including probabilistic methods for clustering and visualization, generative models of affinity matrices, and approximate inference.

Research Intern

1999

Mitsubishi Electric Research Laboratories (MERL). Cambridge, MA

Analyzed probability models with entropic priors. Developed and implemented algorithms for human body pose estimation and tracking using Hidden Markov Models.

Research/Teaching Assistant

1997-2001

Computer Science Research Lab. Boston University. Boston, MA

Developed and implemented algorithms for learning probabilistic models from rendering synthetic computer graphics data, 3D articulated pose inference, visual tracking of humans, and integration of generative and discriminative models. Coordinated undergraduate project on 3D body pose synthesis. Carried out recitations, grading, and student tutoring for various computer vision and graphics classes. Topics included: shape classification, face recognition and detection, image segmentation, physics-based computer animation and rendering, texture mapping, rendering 3D models, image mosaics, etc. Other: Theory of Compilers, Introduction to C++, Introductory and Advanced Data Structures (undergraduate level).

Software Engineer

1994-1996

Industrial Development and Research Center. Venezuela

Software design and development for financial and retail business applications (C/C++)

Awards and Fellowships

Distinguished Paper Award, <i>International Joint Conference on Artificial Intelligence</i> .	2007
University of Toronto. Elect./Comp. Engineering Postdoctoral Fellowship, Canada.	2002-2003
<i>Neural Information Processing Systems</i> (NIPS) Travel Award. Denver, CO, USA.	2000
Computer Science Research Scholarship, Boston University. Boston, MA, USA.	1998-2001
UCLA-CONICIT Graduate Fellowship Award for Ph.D. in Computer Science.	1996-2000
<i>José Félix Rivas</i> National Medal (III class). National Merit Award. Venezuela.	1996
<i>Magna cum laude</i> , first in graduating class (ranked 1/123). Informatics Engineering.	1995
Universidad Centroccidental (UCLA), Venezuela.	

Publications in Proceedings/Journals (peer reviewed)

1. Y. Yan, R. Rosales, G. Fung, M. Schmidt, L. Bogoni, G. Hermosillo, L. Moy, and J. Dy. **Modeling Annotator Expertise: Learning when Everybody Knows a Bit of Something.** To Appear in Proc. *International Conference on Artificial Intelligence and Statistics (AISTATS)* 2010.
2. V. Vural, G. Fung, R. Rosales, and J. Dy. **Multi-Class Classifiers and Their Underlying Shared Structure.** In Proc. *International Joint Conference on Artificial Intelligence (IJCAI)* 2009.
3. S. Yu, B. Krishnapuram, R. Rosales, and R. B. Rao. **Active Sensing.** In Proc. *International Conference on Artificial Intelligence and Statistics (AISTATS)* 2009.
4. M. Müller, R. Rosales, H. Steck, S. Krishnan, B. Rao, and S. Kramer. **Subgroup Discovery for Test Selection: A Novel Approach and its Application to Breast Cancer Diagnosis.** In Proc. *Intelligent Data Analysis (IDA)* 2009.
5. M. Müller, R. Rosales, H. Steck, S. Krishnan, B. Rao, and S. Kramer. **Data-Efficient Information-Theoretic Test Selection.** In Proc. *Conference on Artificial Intelligence in Medicine (AIME)* 2009.
6. W. Tong, R. Rosales, and G. Fung, and. **Automatic Discrimination of Mislabeled Training Points for Large Margin Classifiers.** (*Snowbird*) *Machine Learning Workshop* 2009.
7. S. Yu, G. Fung, R. Rosales, and R. B. Rao. **Privacy-Preserving Cox Regression for Survival Analysis.** In Proc. *Knowledge Discovery and Data Mining (KDD)* 2008.
8. M. Schmidt, K. Murphy, G. Fung, and R. Rosales. **Discriminative Structure Learning in Random Fields and an Application for Heart Motion Abnormality Detection.** In Proc. *Computer Vision and Pattern Recognition (CVPR)* 2008.
9. R. B. Rao, G. Fung, and R. Rosales. **On the Dangers of Cross-Validation.** An Experimental Evaluation. In Proc. *SIAM Data Mining (SDM)* 2008.
10. S. Yu, B. Krishnapuram, R. Rosales, H. Steck, and R. B. Rao. **Bayesian Multi-View Learning.** In Proc. *Neural Information Processing Systems (NIPS)* 2007.
11. O. Yakhnenko, L. Lita, R. Rosales, and S. Niculescu. **Principled Generative-Discriminative Hybrid Hidden Markov Model.** In *Neural Information processing Systems, Workshop on Representations and Inference on Probability Distributions (NIPS)* 2007.
12. R. Rosales, P. Krishnamurthy, and R. B. Rao. **Semi-supervised Active Learning for Modeling Medical Concepts from Free Text.** In Proc. *International Conference on Machine Learning Applications (ICMLA)*, 2007.
13. M. Schmidt, G. Fung, and R. Rosales. **Fast Optimization Methods for ℓ_1 Regularization: A Comparative Study and Two New Approaches.** In Proc. *European Conference on Machine Learning (ECML)* 2007.
14. G. Fung, R. Rosales, and R. B. Rao. **Feature Selection and Kernel Design via Linear Programming.** In Proc. *International Joint Conference on Artificial Intelligence (IJCAI)* 2007.
15. Maleeha Qazi, G. Fung, S. Krishnan, R. Rosales, H. Steck, R. B. Rao, Don Poldermans, and Dhanalakshmi Chandrasekaran. **Automated HeartWall Motion Abnormality Detection from Ultrasound Images using Bayesian Networks.** In Proc. *International Joint Conference on Artificial Intelligence (IJCAI)* 2007. [Distinguished Paper Award]
16. T. Wilmes, K. Bohy, A. Gilson, R. Rosales, S. Krishnan, S. Niculescu, M. Qazi, F. Rahmanian, W. Landi, B. Rao. **Automated Chart Abstraction Can Provide Highly Accurate Data Extraction For Clinical Quality Measures: Assessment of REMIND for CMS Heart Failure Measures.** *Circulation* 114: II_864-a. American Heart Association 2006.
17. R. Rosales and G. Fung. **Learning 'Sparse' Metrics via Linear Programming.** In Proc. *Knowledge Discovery and Data Mining (KDD)* 2006.
18. R. Rosales and S. Sclaroff. **Combining generative and discriminative models in a framework for articulated pose estimation.** *International Journal of Computer Vision (IJCV)*. Volume 67 (3) p. 251-276, 2006.

19. G. Fung, R. Rosales, and B. Krishnapuram. **Learning Rankings via Convex Hull Separation**. In Proc. *Neural Information Processing Systems (NIPS)* 2005.
20. V. Gottipaty, R. Rosales, P. Aloni, John Beard, Paul Zimmermann, Linda Adams, R. B. Rao. **Automated Identification of MADIT-II Eligible Patients Using REMIND Artificial Intelligence Software**. *Circulation* 111:e310-e359. American Heart Association 2005.
21. R. Rosales and T. Jaakkola. **Focused Inference**. In Proc. *International Workshop on Artificial Intelligence and Statistics (AISTATS)* 2005.
22. R. Rosales, K. Achan, and B. Frey. **Learning to Cluster using Local Data Topology**. In Proc. *International Conference in Machine Learning (ICML)*, 2004.
23. R. Rosales, K. Achan, and B. Frey. **Unsupervised Image Translation**. In Proc. *International Conference on Computer Vision (ICCV)*, 2003.
24. R. Rosales, K. Achan, and B. Frey. **Translating Images by Unsupervised Estimation of Switching Filters**. Invited paper. In Proc. *Statistical Signal Processing (SSP)*, 2003.
25. R. Rosales and B. Frey. **Learning Generative Models of Affinity Matrices**. In Proc. *19th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2003.
26. R. Rosales and S. Sclaroff. **A Framework for Heading-Guided Recognition of Human Activity**. *Computer Vision and Image Understanding (CVIU)*, 2003.
27. R. Rosales and S. Sclaroff. **Algorithms for Inference in the Specialized Mappings Architecture 2002**. In Proc. *International Conference on Automatic Face and Gesture Recognition (FG2002)*, 2002.
28. R. Rosales and S. Sclaroff. **Learning Body Pose via Specialized Maps**. In Proc. *Neural Information Processing Systems 14*, 2001.
29. R. Rosales, M. Sidiqqi, J. Alon, and S. Sclaroff. **3D Body Pose through Virtual Cameras**. In Proc. *Computer Vision and Pattern Recognition (CVPR)*, 2001.
30. S. Sclaroff, G. Kollios, M. Betke, and R. Rosales. **Motion Mining**. Proc. 2nd International Workshop on Multimedia Databases and Image Communication, 2001.
31. R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff. **3D Hand Pose Reconstruction using Specialized Mappings**. In Proc. *International Conference on Computer Vision (ICCV)*, 2001.
32. R. Rosales and S. Sclaroff. **Specialized Mappings and the Estimation of Human Body Pose from a Single Image**. In Proc. *Workshop on Human Motion*, 2000.
33. R. Rosales and S. Sclaroff. **Inferring Body Pose without Tracking Body Parts**. In Proc. *Computer Vision and Pattern Recognition (CVPR)*, 2000.
34. R. Rosales and S. Sclaroff. **Learning and Synthesizing Human Body Pose and Motion**. In Proc. *International Conference on Automatic Face and Gesture Recognition (FG2000)*, 2000.
35. R. Rosales and S. Sclaroff. **Trajectory Guided Recognition of Actions**. In Proc. *SPIE.*, 1999.
36. R. Rosales and S. Sclaroff. **3D Trajectory Recovery for Tracking Multiple Objects and Trajectory Guided Recognition of Actions**. In Proc. *Computer Vision and Pattern Recognition (CVPR)*, 1999.
37. R. Rosales and S. Sclaroff. **Improved Tracking of Multiple Humans with Trajectory Prediction and Occlusion Modeling**. In Proc. *Workshop on the Interpretation of Visual Motion (CPVR)*, 1998.

Papers under Review

38. M. Schmidt, G. Fung, and R. Rosales. **Fast ℓ_1 Regularization. Current and New Optimization Algorithms**. *Journal of Machine Learning Research (JMLR)*.
39. S. Yu, B. Krishnapuram, Romer Rosales, H. Steck, and R. B. Rao. **Bayesian Multi-view Learning**. *Transactions Pattern Analysis and Machine Intelligence (PAMI)*.

Technology and Product Demonstrations

40. R. B. Rao, R. Rosales, B. Krishnapuram, S. Krishnan, L. Bogoni, X. Zhou. **Mining Medical Records for Computer Aided Diagnosis**. In *Knowledge Discovery and Data Mining (KDD)*, 2006.
41. Glen. Fung, R. Rosales, B. Krishnapuram, R. B. Rao. **Machine Learning and Computer Aided Diagnosis**. In *International Conference on Machine Learning (ICML)*, 2006.

Patents (Granted/Pending)

- P1. A. Pandya, R. Rosales, B. Rao, H. Steck. **Medical Ontologies for Computer Assisted Clinical Decision Support** [Granted 2009]
- P2. R. Rosales, S. Chung, F. Farooq, G. Fung, B. Krishnapuram, B. Rao, J. W.s, S. Yu. **Automated Patient/Document Identification and Categorization For Medical Data** [Provisional 2009]
- P3. G. Fung, R. Rosales, J. Dy, and Y. Yan. **Automatic Labeler Assignment** [Provisional 2009]
- P4. R. Rosales, M. Bundschus, B. Krishnapuram, F. Rahmanian, B. Rao, S. Yu. **System and Method for Scoring and Ranking Quality Improvement Factors in Patient Care** [Pending 2009]
- P5. R. Rosales. **Method for Automatic Labeling Of Unstructured Data Fragments from Electronic Medical Records** [Pending 2009]
- P6. S. Yu, B. Krishnapuram, R. Rosales, B. Rao. **Active Multi-Modality Sensing for Training Classifiers** [Provisional 2009]
- P7. O. Yakhnenko, R. Rosales, S. Niculescu, L. Lita. **System and Method for Text Tagging and Segmentation Using a Generative/Discriminative Hybrid Hidden Markov Model** [Pending 2008]
- P8. R. Rosales, S. Niculescu, W. Landi, B. Rao, P. Giang, S. Krishnan. **Patient Data De-Identification by Obfuscation** [Pending 2008]
- P9. R. Rosales, G. Fung, M. Schmidt, S. Krishnan, B. Rao. **Learning Classifiers for Computer-Aided Diagnosis using Multiple-Labeler Data Analysis** [Pending 2008]
- P10. R. Rosales, P. Krishnamurthy, B. Rao, H. Steck. **Learning or Inferring Medical Concepts from Medical Transcripts** [Pending 2008]
- P11. M. Yetisgen-Yildiz, S. Niculescu, R. Rosales, B. Rao, S. Krishnan. **Automated Interpretation and Replacement of Date References in Unstructured Text** [Pending 2007]
- P12. B. Rao, S. Krishnan, W. Landi, R. Rosales, S. Niculescu, F. Rahmanian, H. Steck. **Quality Metric Extraction and Editing for Medical Data** [Pending 2007]
- P13. R. Rosales, M. Müller, S. Krishnan, B. Rao. **Guiding Differential Diagnosis through Information Maximization** [Pending 2006]

Software Engineering

Languages: Assembly (x86 and VAX), C/C++, HTML, Java (some experience), Matlab, OpenGL, SQL, XML.

Software Development: hands-on experience with software development processes as manager and developer, Waterfall and Agile methodologies, and Object-Oriented programming.

Other

Spoken languages: fluent in Spanish

Service: founding board member and communications liaison for Siemens Employee Resource Group

Current Professional Service

Program committee/reviewer for the following (recent) conferences: International Conference on Artificial Intelligence and Statistics (AISTATS) 2009, Computer Vision and Pattern Recognition (CVPR) 2010-2007, Neural Information Processing Systems (NIPS) 2008-2007, Bioinformatics and Biomedicine 2009-2008 (BIBM), International Conference on Machine Learning (ICML) 2008, European Conference on Computer Vision (ECCV) 2008, Knowledge Discovery and Data Mining (KDD) 2007

Reviewer for the following (major) journals: Machine Learning Journal, Transactions on Pattern Analysis and Machine Intelligence, Transactions on Image Processing, Neurocomputing, Transactions on Neural Networks, Transactions on Signal Processing, Data Mining and Knowledge Discovery, International Journal of Computer Vision, Pattern Recognition Letters

Guest Editor: Journal of Data Mining and Knowledge Discovery, Special Issue Mining Medical Data (2010)