

Rajesh G. Parekh

TEL: 650.793.0243

e-mail: rparekh AT iee DOT org

URL: <http://www.cs.iastate.edu/~parekh>

Interests

Knowledge Discovery and Data Mining in very large scale data, Applied Machine Learning, Web Search and Advertising, Web Mining, Social Network Analysis, Neural Networks, Grammatical Inference.

Education

- **Ph.D. Computer Science** May 1998
[Iowa State University](#), Ames, IA 50011.
Dissertation: [Constructive Learning: Inducing Grammars and Neural Networks](#).
Advisor: [Dr. Vasant Honavar](#).
- **M.S. Computer Science** Dec. 1993
[Iowa State University](#), Ames, IA 50011.
Project: Efficient Learning of Regular Languages.
- **B.E. Computer Technology** Jun. 1991
[Victoria Jubilee Technical Institute](#), Bombay (Mumbai), India.
Project: *Abhas* - Simulator for Parallel Architectures and Computing.

Professional Experience

- **Director, Data Mining and Research, Yahoo!**, since Jun. 2007
Strategic thought leader and manager of a talented data mining team applying large scale data mining and research at Yahoo! to solve challenging business problems such as enhancing user experience, increasing ad monetization, online fraud detection, optimization of large scale display and search engine marketing (SEM) campaigns, user behavior modeling, and social networks analysis.
- **Data Mining Engineering Manager, Yahoo!**, Apr. 2006 – May 2007
Managed a talented team of data miners in the US and in India to deliver applied data mining and machine learning based solutions for key problems at Yahoo!
 - Designed and implemented models for predicting ‘news worthiness’ of queries. These models are being leveraged in multiple applications.
 - Conducted an in-depth study of the rich social networks at Yahoo! and identified strategic applications in ad targeting and content recommendation.
 - Built a framework for keyword recommendations for search engine advertising using collaborative filtering and association rules mining.
 - Designed and implemented a system for personalizing ads based on historical user behavior and preferences. The solution suitably balances user experience with ad monetization.
 - Delivered a machine learning based solution for online fraud detection. This solution resulted in substantial revenue savings.
- **Technical Yahoo! Principal, Yahoo!**, Oct. 2004 – Mar 2006
Led research and analysis efforts for a variety of Data Mining and Research projects at Yahoo!

- Designed and implemented an algorithm for optimizing real-estate allocation on web pages. This algorithm demonstrated a significant improvement in user experience while simultaneously increasing ad revenue.
- Analyzed online user behavior and navigation patterns and built models to predict propensity to purchase high margin items such as automobiles, travel, and high-end electronics.
- Implemented scalable machine learning based approaches to create and optimize large search engine marketing (SEM) campaigns.
- **Senior Data Mining Engineer, Blue Martini Software**, Jun. 2000 – Oct. 2004.
Responsible for the design and implementation of several key components of the Blue Martini Business Intelligence suite, mining client and prospect data to uncover interesting nuggets of information, and providing data mining expertise to Blue Martini sales and professional services.
- **Senior Research Associate**, Allstate Research and Planning Center, Feb. 1998 - Jun. 2000
Applied research in data mining and knowledge discovery with emphasis on insurance problems such as cross-sell, retention, fraud detection, and customer lifetime value modeling.
- **Research Assistant, Iowa State University**, Jan. 1995 - May 1997, summer 1992.
Designed a variety of constructive neural network algorithms with provable convergence properties. Implemented these algorithms in a modular neural network toolbox that is publicly available. This project was funded by the NSF.
- **Instructor, Iowa State University**, summer 1993, 1994, and 1995
Instructor for Computer Literacy and Applications.
- **Teaching Assistant, Iowa State University**, 1992 - 1994, and 1997
Head Teaching Assistant for Computer Literacy and Applications and Teaching Assistant for a variety of senior undergraduate and graduate courses.

Honors

- **Yahoo! Data Wizard Award**, 2005.
- **KDD Cup Honorable Mention**, 2001.
- Paper entitled *Mining Operational Databases to Predict Short-Term Defection Among Insured Households* was cited in the list of **Top 10 Insights** at the Advanced Research Techniques (ART) Forum, 2000.
- **Research Excellence Award**, Iowa State University, 1998.
- **John Vincent Atanasoff Graduate Award**, Iowa State University, 1996.
- **Teaching Excellence Award**, Iowa State University, 1993.
- **Graduate College Scholarship**, Iowa State University, 1992 - 1997
- Member of the **Phi Kappa Phi** and **Upsilon Pi Epsilon** honor societies.
- **Premium for Academic Excellence Award**, Iowa State University, 1991 - 1992.
- **First prize** at IEEE (Bombay section) student paper contest for the paper: *Task allocation model for Distributed Processing Systems*, 1991.
- **J.N.Tata and R.D.Sethna scholarships** for higher education, 1991.
- **Merit Certificate** - 11th out of 77,752 students in H.S.C. (grade XII), 1987.

Professional Activities

- **Program Committee Member**
 - ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD'07**)
 - IEEE International Conference on Data Mining (**ICDM'07**)
 - ICML Workshop on Challenges and Applications of Grammatical Inference (**CAGI'07**)

- International Colloquium on Grammatical Inference conference ([ICGI'98](#), [ICGI'00](#), [ICGI'02](#), [ICGI'04](#), [ICGI'06](#), [ICGI'08](#))
- International Workshop on Business Intelligence for the Real-Time Enterprise ([BIRTE'06](#))
- International Conference on Intelligent Systems, Design, and Applications ([ISDA'03](#))
- International Workshop on Extraction of Knowledge from Databases ([EKDB'03](#))
- Conference on Computational Biology and Genome Informatics ([CBGI'02](#))
- **Reviewer**
 - [Neural Processing Letters](#)
 - [Data Mining and Knowledge Discovery](#)
 - [Machine Learning Journal](#)
 - [IEEE Transactions on Neural Networks](#)
 - [Pattern Recognition Journal](#)
 - [NASA Intelligence Systems Project](#)
 - [Journal of Machine Learning Research](#)
 - [Information Fusion Journal](#)
 - [Asian Journal of Control](#)
 - [ACM SIGKDD KDD'01](#) and [KDD'03](#)
 - International Conference on Machine Learning ([ICML'01](#))
 - ACM Conference on Electronic Commerce ([EC'01](#))
 - SIAM Conference on Data Mining ([SDM'01](#) and [SDM'02](#))
 - [Wiley Scientific Publishing](#)
 - [Irwin Publishing](#)
- **Other Activities**
 - Treasurer, ACM SIGKDD [KDD'04](#)
 - Steering committee member, International Colloquium on Grammatical Inference (1996 - 2002).
 - Web master, [International Grammatical Inference Community homepages](#) (1998 - 2002).
 - Member, Computer Science colloquium committee at Iowa State University (Jan. 1995 - Dec. 1997).
 - *Cy-Tag* mentor for talented high school students (summer 1992 and 1993).
- **Society Memberships**
 - Institute of Electrical and Electronic Engineers (IEEE)
 - ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD).

Publications

Journal

1. Kohavi, R., Mason, L., Parekh, R., and Zheng, Z., [Lessons and Challenges from Mining Retail E-Commerce Data](#). *Machine Learning* vol. 57 (1-2), pp. 83-113, 2004.
2. Parekh, R. and Honavar, V., [Learning DFA from Simple Examples](#). *Machine Learning* vol 44, pp. 9-35, 2001.
3. Parekh, R., and Yang, J., and Honavar, V. [Constructive Neural Network Learning Algorithms for Pattern Classification](#). *IEEE Transactions on Neural Networks*. 11(2), pp. 436-451. 2000.
4. Parekh, R., Yang, J., and Honavar, V., Comparison of Performance of Variants of Single-Layer Perceptron Algorithms on Non-separable Datasets. *Neural, Parallel, and Scientific Computation*, vol. 8, pp. 415-438, 2000.
5. Yang, J., Parekh, R., and Honavar, V., [DistAl: An Inter-Pattern Distance-Based Constructive Learning Algorithm](#). *Intelligent Data Analysis* 3, pp. 55-73. 1999.

Invited Chapters

1. Parekh, R. and Honavar, V., [Automata Induction, Grammar Inference, and Language Acquisition](#), *The Handbook of Natural Language Processing*. Dale, Moisl, and Somers (editors), Marcel Dekker Inc., New York, 2000. pp. 727-764.
2. Honavar, V., Yang, J., and Parekh, R., Constructive Learning and Structural Learning. *Encyclopedia of Electrical and Electronics Engineering*. Webster, J. (editor), New York: Wiley. 1999.

Refereed Conferences and Workshops

1. Kohavi, R., and Parekh, R., [Visualizing RFM Segmentation](#). *Proceedings of the 2004 SIAM International Conference on Data Mining (SDM'2004)*, Orlando, FL, pp. 391-399, April 22-24, 2004.
2. Kohavi, R., and Parekh, R., [Ten Supplementary Analyses to Improve E-Commerce Websites](#). *Proceedings of Webmining as a Premise to Effective and Intelligent Web Applications (WEBKDD'2003)*, Washington D.C., Aug. 27, 2003.
3. Parekh, R., and Honavar, V. [On the Relationship between Models for Learning in Helpful Environments](#). *Proceedings of the Fifth International Colloquium on Grammatical Inference (ICGI'2000)*, Lisbon, Portugal, pp. 207-220, Sep. 11-13, 2000.
4. Tuason, N., and Parekh, R. [Mining Operational Databases to Predict Short-Term Defection Among Insured Households](#). *Proceedings of the 2000 Advanced Research Techniques Forum (ART'2000)*, Monterey, CA. Jun 4-7, 2000.
5. Parekh, R., and Honavar, V. [Simple DFA are Polynomially Probably Exactly Learnable from Simple Examples](#). *Proceedings of the Sixteenth International Conference on Machine Learning (ICML'99)*, Bled, Slovenia, pp. 298-306, Jun. 27-30, 1999.
6. Yang, J., Parekh, R., Honavar, V., and Dobbs, D. [Data-Driven Theory Refinement Using KBDistAl](#). *Proceedings of the Third International Symposium on Intelligent Data Analysis (IDA'99)*. Amsterdam, The Netherlands, pp. 331-342, 1999.
7. Yang, J., Parekh, R., Honavar, V., and Dobbs, D. [Data-Driven Theory Refinement Algorithms for Bioinformatics](#). *Proceedings of the International Joint Conference on Neural Networks (IJCNN'99)*, Washington, D.C., vol. 6. pp. 4064-4068, 1999.
8. Parekh, R., Nichitiu, C., and Honavar, V., [A Polynomial Time Incremental Algorithm for Learning DFA](#). *Proceedings of the Fourth International Colloquium on Grammatical Inference (ICGI'98)*, Ames, IA. *Lecture Notes in Computer Science* vol. 1433 pp. 37-49, 1998.
9. Parekh, R., Yang, J., and Honavar, V., [Constructive Theory Refinement in Knowledge Based Neural Networks](#). *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks (IJCNN'98)*, Anchorage, AK, pp. 2318-2323, May 4-8, 1998.
10. Yang, J., Parekh, R., and Honavar, V., [DistAl: An Inter-Pattern Distance-Based Constructive Learning Algorithm](#). *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks (IJCNN'98)*, Anchorage, AK, pp. 2208-2213, May 4-8, 1998.
11. Parekh, R., and Honavar, V., [Learning DFA from Simple Examples](#). *Proceedings of the Eighth International Workshop on Algorithmic Learning Theory (ALT'97)*, Sendai, Japan. Oct. 6-8, 1997. *Lecture Notes in Computer Science* vol. 1316, pp. 116-131, 1997. A version of this paper was presented at the *Workshop on Grammar Inference, Automata Induction, and Language Acquisition (ICML'97)*, Nashville, TN. Jul. 12, 1997.
12. Parekh, R., Yang, J., and Honavar, V., [Pruning Strategies for the Mtiling Constructive Learning Algorithm](#). *Proceedings of the IEEE/INNS International Conference on Neural Networks (ICNN'97)*, Houston, TX, vol III, pp. 1960-1965, June 9-12, 1997.
13. Parekh, R., Yang, J., and Honavar, V., [MUpstart - A Constructive Neural Network Learning Algorithm for Multi-Category Pattern Classification](#). *Proceedings of the IEEE/INNS International Conference on Neural Networks (ICNN'97)*, Houston, TX, vol III, pp. 1924-1929, June 9-12, 1997.
14. Parekh, R., and Honavar, V., [An Incremental Interactive Algorithm for Regular Grammar Inference](#). *Proceedings of the Third International Colloquium on Grammatical Inference (ICGI'96)*, Montpellier, France. *Lecture Notes in Computer Science* vol. 1147, pp. 238-250, September 24-27, 1996.

15. Yang, J., Parekh, R., and Honavar, V., [MTiling - A Constructive Neural Network Learning Algorithm for Multi-Category Pattern Classification](#). *Proceedings of the WCNN'96*, San Diego, CA, pp. 182-187, September 15-20, 1996.
16. Racherla, G., Killian, S., Fife, L., Lehmann, M., and Parekh, R., [PARSIT: A parallel algorithm reconfiguration simulation tool](#). *Proceedings of the 1995 International Conference on High Performance Computing*, New Delhi, India, pp. 373-378, December 27-30, 1995.
17. Chen, C-H., Parekh, R., Yang, J., Balakrishnan, K., and Honavar, V., [Analysis of Decision Boundaries Generated by Constructive Neural Learning Algorithms](#). *Proceedings of the World Congress on Neural Networks (WCNN'95)*, Washington D.C., pp. 628-635, July 17-21, 1995.
18. Parekh, R., and Honavar, V., Efficient Learning of Regular Languages using Teacher Supplied Positive Samples and Learner Generated Queries. *Proceedings of the Fifth UNB Conference on AI*, Fredrickton, Canada, pp. 195-203, August 11-14, 1993.
19. Parekh R., Balakrishnan, K., and Honavar, V., [An Empirical Comparison of Flat-Spot Elimination Techniques in Back Propagation Neural Networks](#). *Proceedings of the SIMTEC/WNN'92*, Clear Lake, TX, pp. 463-468, November 4-6, 1992.
20. Arunkumar, S., Lal, R., Venkatagopal, R., Parekh, R. and Daruwala, R., SIMPAC: A Simulator for Parallel Architectures and Computing. *Proceedings of the IEEE Region 10 Conference, Tencon'92* Melbourne, Australia, pp. 196-200, 1992.

Ph.D. Dissertation

Parekh, R. [Constructive Learning: Inducing Grammars and Neural Networks](#). Iowa State University, Ames, IA, 1998.

Patents

1. Parekh, R., Parikh, J., and Berkhin, P. Predicting Newsworthy Queries using Combined Online and Offline Models, Filed April 2008.
2. Parikh, J., Bhamidipati, N., and Parekh, R., Identifying Super-Phrases of Text Strings, Filed April 2008.
3. Parekh, R., Parmar, J. and Berkhin, P. Personalizing Sponsored Search Advertising Layout using User Behavior History, Filed March 2008.
4. Bhatt, R., Parikh, J., Parekh, R., and Berkhin, P. Differentiated Treatment of Sponsored Search Results Based on Search Context. Filed October 2007.
5. Tomkins, A., Ojeh, B., Berkhin, P., Parekh, R., Gaffney, S., Fayyad, U. Bid Optimization in Search Engine Marketing. Filed December 2006.
6. Klots, B., Eneva, E., Mayoraz, N., Berkhin, P., Parekh, R., Chow, R. Evaluating the Performance of Binary Classification Systems. Filed December 2006.
7. Fayyad, U., Berkhin, P., Tomkins, A., Parekh, R., Parikh, J., Sculley, D. Keyword Set and Target Audience Profile Generalization Techniques. Filed May 2006.

Work Eligibility

US Citizen

References

Available on request.