

Machine Learning Applied to Uncovering Gene Regulation

Mark Craven

Abstract

Now that the complete genomes of numerous organisms have been ascertained, key problems in molecular biology include determining the functions of the genes in each organism, the relationships that exist among these genes, and the regulatory mechanisms that control their operation. These problems can be partially addressed by using machine learning methods to induce predictive models from available data. My group is applying and developing machine learning methods for several tasks that involve characterizing gene regulation. In one project, for example, we are using machine learning methods to identify transcriptional control elements such as promoters, terminators and operons. In another project, we are using learning methods to identify and characterize sets of genes that are affected by tumor promoters in mammals.

Our approach to these tasks involves learning multiple models for inter-related tasks, and applying learning algorithms to rich and diverse data sources including sequence data, microarray data, and text from the scientific literature.

Curriculum Vitae

Mark W. Craven, Ph.D.

Home Address:

1705 Vilas Avenue

Madison, Wisconsin 53711

Phone: (608) 280-0952

Office Address:

Department of Biostatistics and Medical Informatics

University of Wisconsin-Madison

5730 Medical Sciences Center

1300 University Avenue

Madison, WI 53706

Phone: (608) 265-6181

Fax: (608) 263-0415

E-mail: craven@biostat.wisc.edu

Homepage: <http://www.cs.wisc.edu/~craven/>

Research Interests

Machine Learning, Bioinformatics, Artificial Intelligence, Data Mining,
Information Extraction from Text.

Education

Ph.D. in Computer Sciences, University of Wisconsin-Madison, 1996

Advisor: Prof. Jude Shavlik

Thesis: Extracting Comprehensible Models from Trained Neural Networks

M.S. in Computer Sciences, University of Wisconsin-Madison, 1991

B.A. in Political Science, University of Colorado at Boulder, 1987

Val B. Fischer Award for outstanding Political Science student

Phi Beta Kappa

Professional Experience

7/99 present Assistant Professor, Department of Biostatistics and Medical Informatics
Department of Computer Sciences, University of Wisconsin

11/99 present Member, University of Wisconsin Comprehensive Cancer Center

10/96 6/99 Postdoctoral Fellow, Department of Computer Science, Carnegie Mellon University

6/90 9/96 Research Assistant, Department of Computer Sciences, University of Wisconsin

9/89 5/90 Teaching Assistant, Department of Computer Sciences, University of Wisconsin

6/87 8/89 Member of Technical Sta, GTE Government Systems Grants

10/00 9/03 Adaptive Information Monitoring and Extraction from NIH (1R01 LM07050-01), \$864,000

Invited Lectures

"Machine Learning Applied to Uncovering Gene Regulation"

Biological Research Information Center Symposium on Bioinformatics, (Pohang, Korea), November 2000; TBA, University of Illinois, November 2000.

"Learning Relational and Probabilistic Models in Sequence-Based Domains"

Workshop on Attribute-Value and Relational Learning: Crossing the Boundaries (held in conjunction with the International Conference on Machine Learning), July 2000.

"Extracting Information from Biomedical Text Sources"

SmithKline Beecham Research, July 2000.

"Learning from Rich Data Sources in Order to Understand Genomes"

Oregon State University, April 2000.

"Combining Relational and Statistical Methods for Information Extraction"

Rutgers University, March 1999; University of Wisconsin, March 1999; University of Colorado, March 1999; Columbia University, April 1999.

"Learning to Extract Symbolic Knowledge from Online Sources"

Oak Ridge National Laboratory, October 1998.

"Learning to Extract Symbolic Knowledge from the World Wide Web"

Siemens Corporate Research, February 1998.

"First Order Learning for Web Mining"

University of California, San Diego, December 1997.

"Understanding Models Learned by Neural Networks"

Workshop on AI Applications in Solar-Terrestrial Physics, Lund, Sweden, July 1997.

"Rule Extraction as Learning"

Workshop on Rule Extraction from Artificial Neural Networks (held in conjunction with the Neural Information Processing Systems Conference), December 1996.

"Extracting Comprehensible Models from Trained Neural Networks"

Jet Propulsion Laboratory, April 1996; Carnegie Mellon University, April 1996; Curagen Corporation, March 1996; SmithKline Beecham Research, March 1996; Silicon Graphics Inc., March 1996; General Electric Research, February 1996.

"Understanding Time-Series Networks: A Case Study in Rule Extraction"

Workshop on Noisy Time Series (held in conjunction with the Neural Information Processing Systems Conference), December 1995.

"Extracting Comprehensible Symbolic Concept Representations from Trained Neural Networks"

University of Aberdeen, August 1995; Oxford University, July 1995; University of Colorado, February 1994.

"Constructive-Induction Challenges in Molecular Biology"

Keynote invited talk at the Constructive Induction Workshop (held in conjunction with the International Conference on Machine Learning), July 1994.

Professional Activities

Co-instructor for tutorial on Strategies for Seeking and Publishing Biomedical Literature on the World Wide Web at the 2000 International Conference on Intelligent Systems for Molecular Biology (ISMB)

Co-organizer of the 1998 AAAI/ICML workshop Machine Learning for Text Categorization Invited participant for panel discussion on "The Information Systems Revolution Hasn't Started Yet", DARPA/Rome Labs Information 2000 Workshop, October, 1997

Consultant for GTE Government Systems and Orincon Inc. on data-mining applications, 1997 present

Program committee member for:

International Conference on Knowledge Capture (2001)

International Conference on Machine Learning (ICML) (1997, 1999, 2000)

Knowledge Discovery & Data Mining Conference (KDD) (2000, 2001)

National Conference on Artificial Intelligence (AAAI) (2000)

ECML workshop on Towards Adaptive NLP-driven Systems (1998)

ICML workshop on Attribute-Value and Relational Learning: Crossing the Boundaries (2000)

KDD workshop on Text Mining (2000)

NIPS workshop on Rule Extraction from Artificial Neural Networks (1996)

WWW-8 workshop on Adaptive Systems and User Modeling on the World Wide Web (1999)

Reviewer for:

AAAI Student Abstract and Poster Program (2000)

ACM Transactions on Software Engineering (1998)
Applied Intelligence (1998)
Artificial Intelligence (1996, 1999, 2000)
Australian Conference on Neural Networks (1998)
Bioinformatics (2000)
Communications of the ACM (1995)
Computational Learning Theory and Natural Learning Systems (1994)
Computer Vision Symposium (1995)
Connection Science (1995)
IEEE Transactions on Computers (1997)
IEEE Transactions on Knowledge and Data Engineering (1998, 1999)
IEEE Transactions on Neural Networks (1996, 1997, 1998, 1999)
International Joint Conference on Artificial Intelligence (IJCAI) (1999, 2001)
Journal of Intelligent Information Systems (1997)
Knowledge and Information Systems (1999)
Knowledge-Based Systems (1995)
Machine Learning (1994, 1996, 1997, 1998, 1999, 2000)
Neural Computation (2000)
Neural Information Processing Systems (NIPS) (1997, 1998, 1999, 2000)
Neural Networks (1997)
Nucleic Acids Research (2000)

Publications

Journal and Magazine Articles

M. Craven and S. Slattery. Relational Learning with Statistical Predicate Invention: Better Models for Hypertext. Accepted to Machine Learning.

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery. Learning to Construct Knowledge Bases from the World Wide Web. Accepted to Artificial Intelligence.

M. W. Craven and J. W. Shavlik (1997). Understanding Time-Series Networks: A Case Study in Rule Extraction. International Journal of Neural Systems (Special Issue on Data Mining in Finance), 8(4).

M. W. Craven and J. W. Shavlik (1997). Using Neural Networks for Data Mining. Future Generation Computer Systems (Special Issue on Data Mining), 13(2-3).

M. W. Craven and J. W. Shavlik (1994). Machine Learning Approaches to Gene Recognition. IEEE Expert, 9(2).

M. W. Craven and J. W. Shavlik (1992). Visualizing Learning and Computation in Artificial Neural Networks. *International Journal on Artificial Intelligence Tools*, 1(3).

Book Chapter

M. W. Craven and J. W. Shavlik (1995). Investigating the Value of a Good Input Representation. In Petsche, T., Hanson, S. and Shavlik, J., editors, *Computational Learning Theory and Natural Learning Systems*, Volume 3, MIT Press.

Rigorously Refereed Conference Papers

M. Craven, D. Page, J. Shavlik, J. Bockhorst, and J. Glasner (2000). A Probabilistic Learning Approach to Whole-Genome Operon Prediction. In *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*.

(Among 29% of submissions accepted.)

M. Craven, D. Page, J. Shavlik, J. Bockhorst, and J. Glasner (2000). Using Multiple Levels of Learning and Diverse Evidence Sources to Uncover Coordinately Controlled Genes. In *Proceedings of the 17th International Conference on Machine Learning*.

(Among 43% of submissions accepted.)

M. Craven and J. Kumlien (1999). Constructing Biological Knowledge Bases by Extracting Information from Text Sources. In *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology*.

(Among 30% of submissions accepted.)

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery (1998). Learning to Extract Symbolic Knowledge from the World Wide Web. In *Proceedings of the 15th National Conference on Artificial Intelligence*.

(Among 30% of submissions accepted.)

S. Slattery and M. Craven (1998). Combining Statistical and Relational Methods for Learning in Hypertext Domains. In *Proceedings of the 8th International Conference on Inductive Logic Programming*.

(Among 12% of submissions accepted for extended presentation.)

M. Craven, S. Slattery and K. Nigam (1997). First Order Learning for Web Mining. In *Proceedings of the 10th European Machine Learning Conference*.

M. W. Craven and J. W. Shavlik (1995). Extracting Tree-Structured Representations of Trained Networks.

In *Advances in Neural Information Processing Systems 8*.

(Among 7% of submissions accepted for oral presentation.)

J. C. Jackson and M. W. Craven (1995). Learning Sparse Perceptrons. In *Advances in Neural Information Processing Systems 8*.

(Among 7% of submissions accepted for oral presentation.)

M. W. Craven, R. J. Mural, L. J. Hauser and E. C. Uberbacher (1995). Predicting Protein Folding Classes without Overly Relying on Homology. In Proceedings of the 3rd International Conference on Intelligent Systems for Molecular Biology.

(Among 30% of submissions accepted for oral presentation.)

M. W. Craven and J. W. Shavlik (1994). Using Sampling and Queries to Extract Rules from Trained Neural Networks. In Proceedings of the 11th International Conference on Machine Learning.

(Among 23% of submissions accepted.)

M. W. Craven and J. W. Shavlik (1993). Learning to Represent Codons: A Challenge Problem for Constructive Induction. In Proceedings of the 13th International Joint Conference on Artificial Intelligence.

(Among 25% of submissions accepted.)

M. W. Craven and J. W. Shavlik (1993). Learning Symbolic Rules Using Artificial Neural Networks. In Proceedings of the 10th International Conference on Machine Learning.

(Among 5% of submissions accepted for plenary presentation.)

Other Conference and Workshop Papers

M. Craven (1999). Learning to Extract Relations from MEDLINE. In Working Notes of the AAAI Workshop on Machine Learning for Information Extraction.

M. Craven (1998). Using Statistical and Relational Methods to Characterize Hyperlink Paths. In Working Notes of the AAAI Fall Symposium on Link Analysis.

S. Slattery and M. Craven (1998). Learning to Exploit Document Relationships and Structure: The Case for Relational Learning on the Web. In Proceedings of the Conference on Automated Learning and Discovery (CONALD).

D. W. Opitz, M. W. Craven and J. W. Shavlik (1997). Using Neural Networks to Automatically Re-encode Expert System Knowledge-Bases: Experiments in the NYNEX MAX Domain. In Proceedings of the IEEE International Conference on Neural Networks.

M. W. Craven and J. W. Shavlik (1995). Extracting Comprehensible Concept Representations from Trained Neural Networks. In Working Notes of the IJCAI Workshop on Machine Learning and Comprehensibility.

M. W. Craven and J. W. Shavlik (1993). Understanding Neural Networks via Rule Extraction and Pruning. In Proceedings of the Connectionist Models Summer School.

M. W. Craven and J. W. Shavlik (1993). Extracting Symbolic Rules from Artificial Neural Networks. In Proceedings of the 2nd International Workshop on Multistrategy Learning.

M. W. Craven and J. W. Shavlik (1993). Learning to Predict Reading Frames in E. coli DNA. In Proceedings of the 26th Hawaii International Conference on System Sciences (Biocomputing Track).

G. G. Towell, M. W. Craven, and J. W. Shavlik (1991). Constructive Induction in Knowledge-Based Neural Networks. In Proceedings of the 8th International Workshop on Machine Learning October, 2000