

Keith Noto

Currently a Postdoctoral Associate at Tufts University
161 College Avenue Medford, MA 02155 · noto@cs.tufts.edu · (608) 695-3215

Education

Postdoc, Tufts University, 2009-Present

Postdoc, University of California at San Diego, 2008

PhD, University of Wisconsin—Madison, Computer Sciences, 2007.

Thesis Title: *Learning Expressive Computational Models of Gene Regulatory Sequences and Responses*

MS, University of Wisconsin—Madison, Computer Sciences, 2002.

BA, University of Wisconsin—Madison, Computer Sciences and Philosophy, 1998.

Selected Peer-Reviewed Publications

K. Noto, C. E. Brodley, and D. Slonim.

Anomaly Detection Using an Ensemble of Feature Models.

Proceedings of the 10th IEEE International Conference on Data Mining (ICDM 2010)

(acceptance rate 19%),

IEEE Computer Society Press.

M. H. Saier, Jr., M. R. Yen, K. Noto, D. G. Tamang, and C. Elkan.

The Transporter Classification Database: Recent Advances.

Nucleic Acids Research 37:(Database issue), D274-D278, 2009.

K. Noto, M. H. Saier, Jr. and C. Elkan.

Learning to Find Relevant Biological Articles Without Negative Training Examples.

Twenty-First Australasian Joint Conference on Artificial Intelligence (AI 2008)

(acceptance rate 29%),

In *Lecture Notes in Bioinformatics* 5360:202-213. Springer-Verlag.

C. Elkan and K. Noto.

Learning Classifiers from Only Positive and Unlabeled Data.

Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and

Data Mining (KDD 2008), 213-220 (acceptance rate 19%).

K. Noto and M. Craven.

Learning Hidden Markov Models for Regression using Path Aggregation.

Proceedings of the 24th Uncertainty in Artificial Intelligence Conference (UAI 2008), 444-451

(acceptance rate 13%).

K. Noto and M. Craven.

Learning Probabilistic Models of *cis*-Regulatory Modules that Represent

Logical and Spatial Aspects.

Proceedings of the 2006 European Conference on Computational Biology

(acceptance rate 18%),

In *Bioinformatics* 23(2):e156-162.

K. Noto and M. Craven.

A Specialized Learner for Inferring Structured *cis*-Regulatory Modules.

BMC Bioinformatics 7:528, 2006.

K. Noto and M. Craven.

Learning Regulatory Network Models that Represent Regulator States and Roles.

RECOMB 2004 Workshop on Regulatory Genomics.

In *Lecture Notes in Bioinformatics* 3318, pages 52-64. Springer-Verlag, 2004.

Presentations and Posters

Paper presentation, 2010 International Conference on Data Mining (ICDM) in Sydney, Australia.

Paper presentation, 2008 Conference on Uncertainty in Artificial Intelligence (UAI) in Helsinki, Finland.

Invited speaker, National Library of Medicine (NLM) Informatics Training Conference 2006, hosted by Vanderbilt University.

Paper presentation, 2006 European Conference on Computational Biology (ECCB) in Eilat, Israel.

K. Noto and M. Craven, Poster: A Specialized Learner for Inferring Structured *cis*-Regulatory Modules. 2005 International Conference on Intelligent Systems for Molecular Biology (ISMB) in Detroit, MI.

Paper presentation, RECOMB 2004 Workshop on Regulatory Genomics, hosted by the University of California—San Diego.

Research Experience

(The following briefly highlights my research interests and experience. Please see my research statement for details.)

Research Interests

- Machine learning algorithm development and application.
- Applying machine learning to biological domains (bioinformatics).
- Development of software beyond research level to create real-world tools.

Research Projects

- Semi- and unsupervised anomaly detection in data sets with a very large number of features; specifically, anomaly detection in human fetal microarray data (ICDM 2010, DMKD *in preparation*).
- Biomedical text classification for a specialized database; specifically, the discovery of relevant biomedical journal articles for a specialized database from a small set of positive instances (KDD 2008, AJCAI 2008, TCBB 2009, NAR 2009).
- Sequence-based regression; specifically, predicting real-valued gene expression measurements from DNA sequences based on the presence and arrangement of DNA binding site motifs (UAI 2008).
- Discovery of transcription factor binding site modules, including spatial and logical relationships between sites, primarily from sequence and expression data (ECCB 2006, BMC Bioinformatics 2006).
- Inference of gene regulatory networks from differential gene expression (microarray) and cellular conditions (RECOMB workshop 2004).

Professional Service

Reviewer, *Machine Learning*, *Journal of Machine Learning Research*, *IBM Journal of Research and Development*, *Journal of Information Science and Engineering*, *Data Mining and Knowledge Discovery*, *BMC Bioinformatics*, *IEEE/ACM Transactions on Systems, Man, and Cybernetics*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Member, AAAI program committee in 2008-2011.

Panelist, NSF *robust intelligence* grant review, 2009.

Teaching Experience

Instructor, *Introduction to Artificial Intelligence* at the University of Wisconsin—Madison, 2002.

Instructor, *Introduction to Computer Science*, at Tufts University (Summer 2010).

Scheduled to teach *Data Structures*, at Tufts University (Summer 2011).

Managed student labs for *Introduction to Computer Science* at Tufts University, Spring 2010.

Prepared and gave guest lectures for

- *Introduction to Algorithms* (University of Wisconsin—Madison).
- *Advanced Artificial Intelligence* (University of Wisconsin—Madison).
- *Advanced Bioinformatics* (University of Wisconsin—Madison).
- *Introduction to Artificial Intelligence* (University of California—San Diego).
- *Computational Biology* (Tufts University).

Co-Mentor of an undergraduate student for the Computational Biology and Biostatistics (CBB) summer research program at the University of Wisconsin—Madison, June-August, 2006.

- Designed a two-month computational biology project for the student.
- Advised on all aspects of the project: Code design, computational experiments, analysis of results and a final report.
- Attended a seminar course for mentors.

Teaching assistant at the University of Wisconsin—Madison.

- *Machine Learning* (Fall 2001, Prof. Jude Shavlik).
- *Introduction to Algorithms* (Spring 2002, Prof. Deborah Joseph).
- *Advanced Artificial Intelligence* (Fall 2002, Prof. David Page).
- *Introduction to Artificial Intelligence* (Fall 2002, Prof. Jude Shavlik).

Software

FRaC (Feature Regression and Classification) is a general-purpose anomaly detection algorithm

- Reference: Noto *et al.*, 2010
- Language: Python (Also Uses: WEKA, LIBSVM)
- URL: bcb.cs.tufts.edu/frac

PMAC (PubMed Article Classifier) is an online system for automatically recommending articles relevant to your specialized biomedical database

- Reference: Noto *et al.*, 2008; Elkan and Noto, 2008
- Language: PHP (Also uses: LIBSVM)
- URL: www.cs.tufts.edu/~noto/pmac

RUPA (Regression Using Path Aggregation) is an algorithm for learning a regression model (based on hidden Markov models) from variable-length training sequences.

- Reference: Noto and Craven, 2008
- Languages: C and Python
- URL: www.cs.tufts.edu/~noto/pub/software/rupa/

SCRM (Structured *cis*-Regulatory Modules) are packages for learning the arrangements of multiple DNA-binding sites from sequence data.

- Reference: Noto and Craven, 2006a; Noto and Craven, 2006b
- Language: C++
- URL: <http://pages.cs.wisc.edu/~noto/crm>

Awards

Computation and Informatics in Biology and Medicine (CIBM) fellowship, April 2003-2006 (<http://www.cibm.wisc.edu>).

CIBM fellowship (second award), April 2006-2007

Travel fellowship award for ECCB 2006 from Biosapiens, funded by the EU.

Travel fellowship award for the RECOMB 2004 workshop.

Distinctive scholastic achievement (top 20% of graduating class), B.A. from the University of Wisconsin—Madison, 1998.

Previous Employment

From September 1998 to January 2002, I was a professional software engineer for the Legislative Technology Services Bureau (Wisconsin state government).

I developed software used by the Wisconsin state representatives, senators and staff, including several features of the Assembly and Senate voting systems, and software that communicates between the voting systems and Wisconsin's assembly and senate "InSession" sites.

References

Prof. Carla Brodley
Tufts University, Dept. of Computer Science
161 College Ave. Tufts University Medford, MA 02155
E-mail: brodley@cs.tufts.edu
Phone: (617) 627-3652

Prof. Mark Craven
University of Wisconsin—Madison, Dept. of Biostatistics and Medical Informatics
1300 University Avenue #6730 MSC, Madison WI 53706
E-mail: craven@biostat.wisc.edu
Phone: (608) 265-6181

Prof. Charles Elkan
University of California at San Diego, Dept. of Computer Science and Engineering
9500 Gilman Drive (EBU3B, MC0404), La Jolla, CA 92093
E-mail: elkan@cs.ucsd.edu
Phone: (858) 534-8897

Prof. Sam Guyer
Tufts University, Dept. of Computer Science
161 College Ave. Tufts University Medford, MA 02155
E-mail: sguyer@cs.tufts.edu
Phone: (617) 627-4948

Prof. Ben Hescott
Tufts University, Dept. of Computer Science
161 College Ave. Tufts University Medford, MA 02155
E-mail: hescott@cs.tufts.edu

Prof. David Page
University of Wisconsin—Madison, Dept. of Biostatistics and Medical Informatics
1300 University Avenue #6743 MSC, Madison WI 53706
E-mail: page@biostat.wisc.edu
Phone: (608) 265-6186

Prof. Milton H. Saier, Jr.
University of California at San Diego, Dept. of Biology
9500 Gilman Drive, La Jolla, CA 92093
E-mail: saier@biomail.ucsd.edu
Phone: (858) 534-4084

References (Continued)

Prof. Jude Shavlik
University of Wisconsin—Madison, Dept. of Computer Sciences
1210 West Dayton Street, Madison WI 53706
E-mail: shavlik@cs.wisc.edu
Phone: (608) 263-7625

Prof. Donna Slonim
Tufts University, Dept. of Computer Science
161 College Ave. Tufts University Medford, MA 02155
E-mail: slonim@cs.tufts.edu
Phone: (617) 627-2225

Jeffrey Ylvisaker
Director, Legislative Technology Services Bureau (Wisconsin)
17 W. Main St. Suite 200, Madison WI 53703
E-mail: jeff.ylvisaker@legis.state.wi.us
Phone: (608) 267-9528