

PADHRAIC SMYTH

Department of Computer Science, Bren Hall 4212
School of Information and Computer Sciences
University of California, Irvine
CA 92697-3435.
telephone: (949) 824 2558
fax: (949) 824 4056
email: smyth@ics.uci.edu

Date of Birth: July 25th 1962

Citizenship: Republic of Ireland, United States (Dual)

Professional Positions

April 1996–present: Professor, Department of Computer Science, University of California, Irvine

- Full Professor: July 2003 to present
- Associate Professor: July 1998 to June 2003
- Assistant Professor: April 1996 to June 1998

January 2007–present: Director, Center for Machine Learning and Intelligent Systems, University of California, Irvine.

October 1988–March 1996: Member of Technical Staff and Technical Group Leader (from 1992), Jet Propulsion Laboratory, California Institute of Technology, Pasadena.

Education

PhD, 1988: California Institute of Technology, Department of Electrical Engineering.

MSEE, 1985: California Institute of Technology, Department of Electrical Engineering.

BE, 1984: National University of Ireland, University College Galway. Bachelor of Engineering (Electronic) with First-Class Honors.

Professional Affiliations

Joint Faculty Appointment with Department of Statistics, UC Irvine, July 2008–present.

Joint Faculty Appointment with Department of Biomedical Engineering, UC Irvine, July 2001–present.

Center for Research on Information Technology and Organizations (CRITO), UC Irvine, Member since 2008.

Institute for Genomics and Bioinformatics, UC Irvine, Member since 2001.

Institute for Mathematical Behavioral Sciences, UC Irvine, Member since 1999.

Founding Director and Founding Executive Committee Board Member of the ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), 1998.

Visiting Principal Researcher, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, 1996–2001.

IEEE Computer, Information Theory and Signal Processing Societies, Member since 1988.

American Statistical Association, Member since 1997.

Association for Computing Machinery (ACM), SIGMOD, SIGIR, SIGKDD, Member since 1999.

Honors and Awards

ACM SIGKDD Innovation Award in 2009.

Best paper awards in 1997 and 2002, and runner-up for best paper awards in 1998 and 2000, at the ACM SIGKDD Knowledge Discovery and Data Mining Conferences. Shortlisted for best paper award at the 2007 ACM/IEEE Joint Conference on Digital Libraries.

IBM Faculty Partnership Award, 2001.

National Science Foundation CAREER award in 1997.

ACM Teaching Award in the Information and Computer Science Department at UC Irvine in the 1996-97 academic year.

1996 NASA Group Achievement award for “outstanding achievements in design, implementation, and deployment of an advanced analysis system for data mining of large image collections.”

Lew Allen Award for Excellence in Research at JPL in 1993.

17 NASA Certificates for Technical Innovation between 1991 and 1996.

Consulting and Business Activities

Consultant to EBay (2009-present); DataAnalytics LLC (2009-present); Latham-Watkins (2008-present); Netflix (2006-present); Topicseek LLC (2005-2008); Yahoo! (2005-2008); Strativa (2005); IET (2004-2005); JWDirect (2001-2004); Credit Sciences (2000-2004); Nokia Research (2000); First Quadrant Financial Services (1998-1999); Smith-Kline Beecham Research (1998); AT&T Research (1996-1998).

Professional Activities

Journals: Associate/Action Editor

Journal of the American Statistical Association, 2002 to 2005.

IEEE Transactions on Knowledge and Data Engineering, 2002 to 2004.

Machine Learning Journal, July 1998 to December 2001.

Machine Learning Journal, guest editor of special issue on probabilistic learning, 1997.

Journals and Book Series: Editorial Board/Advisory Board Member

Journal of Machine Learning Research, 2000 to present.

Journal of Data Mining and Knowledge Discovery, 1997 to present.

Chapman and Hall: Series in Computer Science and Data Analysis, 2002 to 2008.

Bayesian Analysis, 2004 to 2007.

Journals: Reviewer

Reviewer for *IEEE Transactions on Information Theory*, *IEEE Trans. on Neural Networks*, *IEEE Trans. on Signal Processing*, *IEEE Trans. on Circuits and Systems*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *IEEE Trans. on Knowledge and Data Engineering*, *Statistics and Computing*, *Journal of Artificial Intelligence Research*, *Pattern Recognition Letters*, *Neural Networks*, *Machine Learning*, *Journal of Machine Learning Research*.

Conference Program Chair Positions

Program Chair for the 2001 Symposium on the Interface between Statistics and Computing, Costa Mesa, CA, June 2001.

General Chair of the Sixth International Workshop on Artificial Intelligence and Statistics, January 1997.

Workshop and Conference Organization Positions

Co-Organizer of Workshop on Scientific Data Mining, Institute for Pure and Applied Mathematics (IPAM), UCLA, January 2002.

Co-Organizer of Workshop on Temporal and Spatial Machine Learning, International Conference on Machine Learning (ICML), 2001.

Panels Chair of the ACM SIGKDD Fifth International Conference on Knowledge Discovery and Data Mining, 1999.

Workshop co-chair of the Massive Datasets workshop at the 1998 Neural Information Processing Conference (NIPS).

Tutorials co-Chair for the 1998 National Conference on Artificial Intelligence, Madison, WI, August 1998.

Tutorials Chair for the ACM SIGKDD Conferences on Knowledge Discovery and Data Mining, 1997 and 1998.

Area Chair for Pattern Recognition in Practice IV and V workshops, 1997 and 1999.

Publicity Chair for the ACM SIGKDD Conferences on Knowledge Discovery and Data Mining, 1995 and 1996.

Conference Reviewing and Program Committees

Neural Information Processing Conference (NIPS): 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2001, 2002, 2003, 2004.

ACM SIGKDD Conference on Knowledge Discovery and Data Mining: 1995 through 2004.

International Workshop on Artificial Intelligence and Statistics: 1993, 1995, 1999, 2001, 2003.

International Conference on Machine Learning (ICML): 1997 and 2001.

Others include the National Conference on Artificial Intelligence (AAAI) in 1996, Induction in Science and Statistics Conference (ISIS) in 1996, International Conference on Pattern Recognition (ICPR) in 1998, the International Joint Conference on Artificial Intelligence (IJCAI), 1995, Uncertainty in AI 2004, and WWW-2009.

Postdoctoral Advisees

Romain Thibaux, 2008-2009, currently at Willow Garage.

Michal Rosen-Zvi, 2003-2004; currently at IBM Research, Israel.

Michael Duff, 2005-2006; currently a researcher at the University of Massachusetts, Amherst.

Alex Ihler, 2005-2006; currently assistant professor, Department of Computer Science, UC Irvine.

Graduate Students

PhD Advisees (Committee Chair)

Chaitanya Chemudugunta, University of California, Irvine, PhD awarded June 2009, currently at Blizzard software.

Seyoung Kim, University of California, Irvine, PhD awarded June 2007, postdoctoral fellow at CMU.

Darya Chudova, University of California, Irvine, PhD, awarded July 2007, currently a senior researcher at Veracyte, Inc.

Sergey Kirshner, University of California, Irvine, PhD awarded March 2005, currently assistant professor, Department of Statistics, Purdue University

Scott Gaffney, University of California, Irvine, PhD awarded March 2004, currently at Yahoo! Research, Sunnyvale, CA.

Xianping Ge, University of California, Irvine, PhD awarded October 2002. Currently at Google, CA.

Igor V. Cadez, University of California, Irvine, PhD awarded May 2002. Currently at Sparta, Laguna Niguel, CA.

Dimitry Pavlov, University of California, Irvine, PhD awarded December 2001. Currently at Yahoo! Research, Sunnyvale, CA.

PhD Students Co-Advised

Kevin Lin, University of California, Irvine, PhD awarded June 2008, postdoctoral fellow at UCSF.

Current PhD Students

Advanced to Candidacy: Arthur Asuncion (2008), Jon Hutchins (2008)

Pre-Candidacy: America Holloway, Drew Frank, Jimmy Foulds, Scott Triglia, Chris DuBois.

PhD Thesis Committee Member

UC Irvine, Computer Science: Vibhav Gogate (2009), Radu Marinescu (2008), Robert Mateescu (2007), Bozhena Bidyuk (2005), Stephen Bay (2001), Irina Rish (1999), Chris Merz (1998), Pedro Domingos (1997).

UC Irvine, Other Departments: Kim Aeling (Microbiology and Molecular Genetics, 2007), Bethany Knapp (Cognitive Science, 2002).

Other Universities: Xuerei Wang (U Mass Amherst, 2009), Sangmin Oh (Georgia Tech, 2009), Carla Domencioni (UC Riverside, 2002), John Lindal (Caltech, 2000), Srinivas Aji (Caltech, 2000), David Babcock (Caltech, 2000), Gavin Horn (Caltech, 1999), Lonnie Chrisman (CMU, 1996), Michael Burl (Caltech, 1996), Barry Ambrose (Caltech, 1995), Zheng Zeng (Caltech, 1995).

PhD Candidacy Committees

UC Irvine, Computer Science: Lars Otten, 2009; Yutian Chen, 2009; Chaitanya Desai, 2008; Pinaki Sinha, 2007; Chloe Azencott, 2007; Guy Yosiphon, 2006; Bo Gong, 2006; Lin Wu, 2005; Yiming Ma, 2004; Dawit Seid, 2004; Suman Sundaresh, 2003; Ye Sun, 2001; Stephen Bay, 1999; Daniel Billsus, 1998; Chris Merz, 1997.

UC Irvine, Other Departments: Tim Rubin, 2009(Cognitive Sciences); John Abatzoglu, 2004 (Earth System Sciences); Mingliang Li, 2002 (Economics); Bethany Knapp, 2000 (Cognitive Sciences); Pei Suen, 1998 (ECE).

Other Universities: Srinivas Aji, 1999 (Caltech); Gavin Horn, 1998 (Caltech); John Lindal, 1998 (Caltech).

Masters Students Supervised

UC Irvine, Information and Computer Science: Vasanth Kumar (2006), Sridevi Parise (2003), Naval Verma (2002), Wagner Truppel (2001), Scott Lundgren (1997).

Royal Institute of Technology, Department of Numerical Analysis and Computing Science, Stockholm: Stefan Edlund, 1997, Thesis entitled *Methods for Cluster Analysis with Applications to Large NASA Data Sets*.

University of Freiburg, Department of Computer Science, MS Diplom Thesis, Daniel Henke, 2007.

Research Grants, Contracts and Gifts

1. *An Information-Theoretic Approach to Distributed Inference and Learning*, from DARPA, AFOSR, and ONR (co-investigator with Rodney Goodman, Electrical Engineering Dept., Caltech).
 - Original award AFOSR-90-0199, February 1st 1990 to May 30th 1992, \$338,161.
 - Continuation award NOO014-92-J-1860: July 1st 1992 to March 30th 1995, \$394,118.
2. *Automated Detection of Natural Features in SAR Images*, Jet Propulsion Laboratory Director's Discretionary Fund, co-investigator with U. M. Fayyad (JPL) and P. Perona (Electrical Engineering Dept., Caltech), January 1st 1994 to December 31st 1994, \$140,000.
3. *Using Information Theory to Discover Patterns in Databases*, Lew Allen Award research grant, Jet Propulsion Laboratory. January 1st 1994 to December 31st 1995, \$25,000, Principal Investigator.
4. *Clustering and Mode Classification of Engineering Time Series Data*, Jet Propulsion Laboratory, June 15th 1996 to October 17th 1996, \$34,401.
5. *A Distributed Biomedical Computing Laboratory*, from NSF (CISE Research Instrumentation), NSF-9617349, co-investigator with L. Bic et al. (University of California, Irvine), March 1 1997 to February 1 1998, \$69,986.
6. *Turbo-Decoding of High Performance Error-Correcting Codes via Belief Propagation*, from AFOSR, grant F49620-97-1-0313, co-investigator with R. J. McEliece (PI, Electrical Engineering Dept., Caltech), May 1 1997 to December 31 1998, \$300,000.
7. *Automated Cloud Screening for Remote Exploration and Experimentation (REE) Applications to the Earth Orbiting-1 (EO-1) Satellite and Similar Platforms*, from the Jet Propulsion Laboratory, June 16th 1997 to November 15th 1997, \$34,601, Principal Investigator.
8. *Exploring QSAR Data using Probabilistic Data Mining*, from SmithKline Beecham Research, July 1st to December 31st 1997, \$35,048, Principal Investigator.
9. *Probabilistic Knowledge Discovery and Data Mining: An Integrated Approach at the Interface of Computer Science and Statistics*, from NSF (CAREER award), NSF-9703120, September 1st 1997 to August 31st 2001, \$304,379, Principal Investigator.
10. *SGER: An Online Repository of Large Data Sets for Data Mining Research and Experimentation*, from NSF, NSF IIS-9813584, co-PIs Dennis Kibler, Michael Pazzani, Aug 15, 1998 to January 31, 2000, \$99,737, Principal Investigator.
11. *Graduate Fellowships in Biomedical Computing*, from US Department of Education, co-investigator with Lubomir Bic, Dennis Kibler, Richard Lathrop, Michael Pazzani, Rodman Shankle, Michael Dillencourt, Richard Granger, Menzel, Gary Lynch. \$750,000. Sept 1, 1997 to August 31, 2001.
12. *Data Mining of High-Dimensional Structure-Activity Data Sets*, from SmithKline Beecham Research, September 1st 1998 to April 1st 1999, \$22,730, Principal Investigator.
13. *Clustering of Sequences and Time Series*, from HNC Software, Inc, \$40,913, January 1 1999 through Dec 31 1999, Principal Investigator.

14. *Data Mining of Multivariate Time-Series Sensor Data for Semiconductor Manufacturing*, from NIST/National Semiconductor corporation, April 1 1999 through Dec 31 2001, \$162,000, Principal Investigator.
15. *Research in Data Mining*, Gift from Microsoft Research, October 1999, \$60,000, Principal Investigator.
16. *Spatio-Temporal Data Mining of Scientific Trajectory Data*, from Lawrence Livermore National Laboratory, March 1st to September 30th 2000, \$42,937, Principal Investigator.
17. *Sequential Data Analysis for Biomedical Applications*, UCI CORCLR Program, July 1 2000 to June 30th 2001, \$12,000, co-investigator with Christine McLaren, UCI Dept of Epidemiology.
18. *Probabilistic Clustering of Dynamic Trajectories for Scientific Data Mining*, Institute for Scientific Computer Research, Lawrence Livermore National Laboratory, October 1 2000 to September 30 2001, \$39,178, Renewal: October 1 2001 to September 30 2002, \$28,448, Principal Investigator.
19. *Data Mining of Digital Behavior*, NSF-IIS-0083489, with Sharad Mehrotra (co-I), September 15th 2001 to August 30th 2004, \$425,000, Principal Investigator.
20. *Predictive Models for Cancer Detection and Therapy*, November 1st 2000 to October 31st 2001, University of California, Irvine, Cancer Research Grants, \$14,301, co-investigator with Christine McLaren, UCI Dept of Epidemiology.
21. *Spatial Data Mining for Massive Scientific Data Sets* from Lawrence Livermore National Laboratory, May 1st 2001 to August 31st 2002, \$100,000, Principal Investigator.
22. *IBM Faculty Partnership Award*, Gift from IBM Watson Research Center, May 18th 2001, \$40,000, Principal Investigator.
23. *Intelligent Time-Series Pattern Matching*, Jet Propulsion Laboratory, June 15th to September 30th 2002, \$80,920, Principal Investigator.
24. *Preclinical Detection and Disease Measurement of Alzheimer's Disease and Related Disorders Using EEG, Psychophysical and Data Mining Methods*, Alzheimer's Association of America, September 1st 2001 to August 30th 2003, \$250,000, co-investigator with Don Hoffmann and Rod Shankle (PI), UCI Department of Cognitive Science.
25. *Predicting Coupled Ocean-Atmosphere Modes With A Climate Modeling Hierarchy*, US Department of Energy: Climate Change Prediction Program, \$396,000, February 1st 2002 to January 31st 2005, co-investigator with Andrew Robertson and Michael Ghil (PI), Department of Atmospheric Sciences, UCLA.
26. *Biomedical Informatics Training Program*, National Institutes of Health and National Library of Medicine, award number LM-07443, \$4,570,204, July 1st 2002 to June 30th 2007, Faculty Mentor.
27. *Entity-Based Data Mining from Spatiotemporal and Text-Based Data Streams*, Knowledge Discovery and Dissemination Program, National Science Foundation, \$1,005,000, October 1st 2002 to June 30th 2006, Principal Investigator.
28. *NSF-ITR: The OptIPuter*, Information Technology Research (ITR) program, National Science Foundation, \$14,000,000, October 1st 2002 to September 30th 2007, co-investigator (PI: Larry Smarr, UCSD).

29. *NSF-ITR: Responding to the Unexpected*, Information Technology Research (ITR) program, National Science Foundation, \$12,500,000, October 1st 2003 to September 30th 2008, co-investigator (PI: Sharad Mehrotra, UC Irvine).
30. *Studies of regional-scale climate variability and change: Hidden Markov models and coupled ocean-atmosphere modes*, funded by the Climate Change Prediction Program, US Department of Energy, October 1st 2004 to September 30th 2007, UCI Principal Investigator.
31. *Statistical Data Mining of Time-Dependent Data with Applications in Geoscience and Biology*, NSF-IIS-0431085, National Science Foundation, \$566,644, October 1st 2004 to September 30th 2007, Principal Investigator.
32. *Transdisciplinary Imaging Genetics Center*, NIH Grant No. 1-P20-RR020837-01, total award is \$1,724,026, 9/28/04 to 7/31/07, co-investigator (PI: Steven Potkin, UCI).
33. *NAMIC: National Alliance for Medical Image Computing, National Institutes of Health*, award number NIH U54 EG005149, total UCI award is \$609,253 from 9/17/04 to 8/31/06, co-investigator (PI: Steven Potkin, UCI).
34. *Morphometry Biomedical Informatics Research Network (MBIRN)*, NIH, total UCI award is \$579,880 from 9/30/04 to 5/31/06, co-investigator (PI: Steven Potkin, UCI).
35. *Bringing Probabilistic Text Mining Techniques to Historical Document Collections: An Early American Case Study*, UCI CORCLR Award MI-05-06-14, \$18,080, 7/1/2005 - 6/30/2006, co-investigator (PI is Sharon Block, History Department, UC Irvine).
36. *UC Irvine Knowledge Discovery Evaluation Challenge Project*, Entity Analytics Division, International Business Machines (IBM), \$73,430, 7/15/05 to 12/31/05, Principal Investigator.
37. *Characterizing ITCZ Dynamics and Breakdown using Statistical Learning Methods and Satellite Data*, National Science Foundation, award number ATM-0530926, \$618,000, 10/1/2005 to 9/30/2008, co-investigator (PI: Gudrun Magnúsdóttir, Department of Earth System Sciences, UC Irvine).
38. *Functional Biomedical Informatics Research Network (FBIRN)*, NIH, \$10,000,000, from Dec 1st 2005 to September 30th 2010, co-investigator (PI: Steven Potkin, UC Irvine).
39. *CRI: Collaborative Research: Improving Experimental Computer Science with a Searchable Web Portal for Datasets*, National Science Foundation, award number CNS-0551510, \$400,000, March 15, 2006 to February 28, 2009, co-PI (with Andrew McCallum, University of Massachusetts).
40. *Collaborative Research: Regional Climate-Change Projections Through Next-Generation Empirical and Dynamical Models*, Department of Energy, Scientific Discovery through Advanced Computing: Climate Change Prediction, award number DE-FG02-07ER64429, \$800,000, Oct 1 2007 to Sept 30 2010, co-PI (with Andrew Robertson, IRI-Columbia).
41. *Scalable methods for the analysis of network-based data*, Office of Naval Research: Multi-disciplinary University Research Initiative (MURI) Award, award number N00014-08-1-1015, \$3,358,578, May 8 2008 to May 7 2011, Principal Investigator.

Publications List

Books

- B1 *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, P. Baldi, P. Frasconi, and P. Smyth, John Wiley, June 2003.
- B2 *Principles of Data Mining*, D. Hand, H. Mannila, and P. Smyth, Cambridge, MA: MIT Press, 2001.
- B3 *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurasamy (eds.), Palo Alto, CA: AAAI/MIT Press, 1996.

Journal Papers

- J1 K. Lin et al, 'Circadian clock genes contribute to the regulation of hair follicle cycling', *PLoS Genetics*, 2009, to appear.
- J2 D. Newman, A. Asuncion, P. Smyth, and M. Welling, 'Distributed algorithms for topic models,' *Journal of Machine Learning Research*, to appear, 2009.
- J3 M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth, and M. Steyvers, 'Learning author-topic models from text corpora,' *ACM Transactions on Information Systems*, to appear, 2009.
- J4 A. Ihler, J. Hutchins, and P. Smyth, 'Learning to detect events with Markov-modulated Poisson processes,' *ACM Transactions on Knowledge Discovery from Data*, 1(3), 2007.
- J5 S. J. Gaffney, A. W. Robertson, P. Smyth, S. J. Camargo and M. Ghil, 'Probabilistic clustering of extratropical cyclones using regression mixture models,' *Climate Dynamics*, 29(4):423–440, 2007
- J6 S. J. Camargo, A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 'Cluster analysis of typhoon tracks. Part I: general properties,' *Journal of Climate*, 20:3635-3653, 2007.
- J7 S. J. Camargo, A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 'Cluster analysis of typhoon tracks. Part II: large-scale circulation and ENSO,' *Journal of Climate*, 20:3654-3676, 2007.
- J8 L. Friedman, Stern, Brown, Mathalon, Turner, Glover, Gollub, Lauriello, Lim, Cannon, Greve, Bockholt, Belger, Mueller, Doty, He, Wells, Smyth, Pieper, Kim, Kubicki, Vangel, and Potkin, 'Test-retest and between-site reliability in a multicenter fMRI study,' *Human Brain Mapping*, 2007.
- J9 A. Ihler, S. Kirshner, P. Smyth, M. Ghil, A. Robertson, 'Graphical models for statistical inference and data assimilation,' *Physica D*, 230(1–2):72–87, 2007.
- J10 S. Kim and P. Smyth, 'Segmental hidden Markov models with random effects for waveform modeling,' *Journal of Machine Learning Research*, 7(Jun):945–969, 2006.
- J11 A. W. Robertson, S. Kirshner, P. Smyth, S. P. Charles, B. Bates, 'Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland,' *Quarterly Journal of the Royal Meteorological Society*, 132, 519–542, 2006.

- J12 Turner, J.A., Smyth, P., Fallon, J.F., Kennedy, J.L., Potkin, S.G., 'Imaging and genetics in schizophrenia,' *Neuroinformatics*, 4(1), 21–50, March 2006.
- J13 A. Robertson, S. Kirshner, and P. Smyth, 'Hidden Markov models for modeling daily rainfall occurrence over Brazil,' *Journal of Climate*, 17(22):4407–4424, November 2004.
- J14 K. K. Lin, D. Chudova, P. Smyth, and B. Andersen, 'Identification of hair cycle-associated genes from time-course gene expression profile data by using replicate variance,' *Proceedings of the National Academy of Sciences*, 101:15955–15960, November 2004.
- J15 D. Pavlov, H. Mannila, and P. Smyth, 'Beyond independence: probabilistic models for query approximation,' *IEEE Transactions on Knowledge and Data Engineering*, 15(6):1409–1421, September 2003.
- J16 I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White, 'Visualization of navigation patterns on a Web site using model-based clustering,' *Journal of Data Mining and Knowledge Discovery*, 7(4), 399–424, 2003.
- J17 D. Chudova and P. Smyth, 'Sequential pattern discovery under a Markov assumption,' *Journal of Data Mining and Knowledge Discovery*, 7(3), 273–299, 2003.
- J18 I. V. Cadez, P. Smyth, G. J. McLachlan, and C. E. McLaren, 'Maximum likelihood estimation of mixture densities for binned and truncated multivariate data,' *Machine Learning*, 47, 7–34, 2002.
- J19 X. Ge, D. Eppstein, and P. Smyth, 'The distribution of cycle lengths in graphical models for iterative decoding' *IEEE Transactions on Information Theory*, v.47, no.6, 2549–2552, September 2001.
- J20 P. Smyth, 'Data mining: data analysis on a grand scale?', *Statistical Methods in Medical Research*, 9, 309–327, 2000.
- J21 P. Smyth 'Model selection for probabilistic clustering using cross-validated likelihood,' *Statistics and Computing*, 9, 63–72, 2000.
- J22 U. Fayyad and P. Smyth, 'Cataloging and mining massive databases for science data analysis,' *Journal of Computational Graphics and Statistics*, 8(3), 589–610, 1999.
- J23 P. Smyth, K. Ide, and M. Ghil, 'Multiple regimes in Northern hemisphere height fields via mixture model clustering,' *Journal of the Atmospheric Sciences*, vol. 56, no. 21, 3704–3723, 1999.
- J24 P. Smyth and D. Wolpert, 'Linearly combining density estimators via stacking,' *Machine Learning*, 36(1/2), 59–83, July 1999.
- J25 M. C. Burl, L. Asker, P. Smyth, U. M. Fayyad, P. Perona, L. Crumpler, and J. Aubele, 'Learning to recognize volcanoes on Venus,' *Machine Learning*, 30(2–3), 165–194, 1998.
- J26 P. Smyth, 'Belief networks, hidden Markov models, and Markov random fields: a unifying view,' *Pattern Recognition Letters*, 18, 1261–1268, 1997.
- J27 C. Glymour, D. Madigan, D. Pregibon, and P. Smyth, 'Statistical themes and lessons for data mining' *Journal of Knowledge Discovery and Data Mining*, 1(1), 11–28, 1997.
- J28 C. Brodley and P. Smyth, 'Applying classification algorithms in practice,' *Statistics and Computing*, 7(1), 45–56, March 1997.

- J29 P. Smyth, D. Heckerman, M. Jordan, 'Probabilistic independence networks for hidden Markov probability models,' *Neural Computation*, 9(2), 227–269, 1997.
- J30 P. Smyth, 'Bounds on the mean classification error rate of multiple experts,' *Pattern Recognition Letters*, 17, 1253–1257, 1996.
- J31 U. M. Fayyad, P. Smyth, N. Weir, and S. Djorgovski, 'Automated analysis and exploration of large image databases: results, progress, and challenges,' *Journal of Intelligent Information Systems*, 4, 7–25, 1995.
- J32 P. Smyth, 'Markov monitoring with unknown states,' *IEEE Journal on Selected Areas in Communications*, special issue on Intelligent Signal Processing for Communications, vol.12, no.9, pp.1600–1612, December 1994.
- J33 A. Y. Lee and P. Smyth, 'Synthesis of minimum-time nonlinear feedback laws for dynamic systems using neural networks,' *Journal of Guidance and Control*, vol.17, no.4, pp.868–870, 1994.
- J34 Z. Zheng, R. Goodman, and P. Smyth, 'Discrete recurrent networks for grammatical inference,' *IEEE Transaction on Neural Networks — Special Issue on Dynamic Recurrent Neural Networks: Theory and Applications*, vol.5, no.2, pp.320–330, March 1994.
- J35 P. Smyth, 'Hidden Markov models for fault detection in dynamic systems,' *Pattern Recognition*, vol.27, no.1, pp.149–164, 1994.
- J36 Z. Zheng, R. Goodman, and P. Smyth, 'Learning finite-state machines with self-clustering recurrent networks,' *Neural Computation*, vol. 5, no. 6, pp.976–990, November 1993.
- J37 J. Miller, R. M. Goodman, and P. Smyth, 'On loss functions which minimize to conditional expected values and posterior probabilities,' *IEEE Transactions on Information Theory*, vol. 39, no.4, pp.1404–1408, July 1993.
- J38 P. Smyth, 'Admissible stochastic complexity models for classification problems,' *Statistics and Computing*, 2, 97–104, 1992.
- J39 R. M. Goodman, P. Smyth, C. Higgins and J. Miller, 'Rule-based networks for classification and probability estimation,' *Neural Computation*, vol. 4, pp.781–804, 1992.
- J40 P. Smyth and R. M. Goodman, 'An information theoretic approach to rule induction from databases,' *IEEE Transactions on Knowledge and Data Engineering*, vol.4, no.4, pp.301–316, August 1992.
- J41 R. M. Goodman and P. Smyth, 'Decision tree design using information theory,' *Knowledge Acquisition*, vol.4, no.1, pp.1–26, 1990.
- J42 R. M. Goodman and P. Smyth, 'Decision tree design from a communication theory standpoint,' *IEEE Transactions on Information Theory*, vol.34, no.5, pp.979–994, September 1988.

Journal Editorials and Commentaries

P. Smyth and S. Kirshner, Commentary on article by T. Ryden, *Bayesian Analysis*, December 2008.

P. Smyth, Commentary on article 'Bump hunting in high-dimensional data,' by Friedman and Fisher, *Statistics and Computing*, 9(2), pp. 149-150, April 1999.

P. Langley, G. M. Provan, P. Smyth, Editorial on probabilistic learning, *Machine Learning Journal*, special issue on probabilistic learning, 91-101, 29 (2/3), November 1997.

Technical Magazine Articles

M1 J. O' Madadhain, J. Hutchins, P. Smyth, Prediction and ranking algorithms for event-based network data, *SIGKDD Explorations* 7(2): 23-30, 2005.

M2 P. Smyth, D. Pregibon, C. Faloutsos, 'Data-driven evolution of data mining algorithms,' *Communications of the ACM*, 45(8), 33-37, August 2002.

M3 C. Apte, B. Liu, E. Pednault, P. Smyth, 'Business applications of data mining,' *Communications of the ACM*, 45(8), 49-53, August 2002.

M4 S. Bay, D. Kibler, M. Pazzani, and P. Smyth, 'The UCI KDD Archive: an online archive of large data sets for data mining research and experimentation,' *ACM SIGKDD Explorations*, 2(2), 81-85, 2000. Also published (in Japanese) in *Information Processing Society of Japan*, IPSJ.

M5 C. Glymour, D. Madigan, D. Pregibon, and P. Smyth, 'Statistical inference and data mining,' *Communications of the ACM*, 39(11), 35-41, November 1996.

M6 U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, 'The KDD process for extracting useful knowledge from volumes of data,' *Communications of the ACM*, 39(11), 27-34, November 1996.

M7 U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, 'From data mining to knowledge discovery,' *AI Magazine*, 37-54, Fall 1996.

M8 G. Piatetsky-Shapiro, C. Matheus, S. Uthurasamy, P. Smyth, 'Knowledge Discovery in Databases (KDD-93): Progress and Remaining Challenges,' *AI Magazine*, pp.77-82, vol.15, no.3, Fall 1994.

Book Chapters

BC1 N. Ashish, R. Eguchi, R. Hegde, C. Huyck, D. Kalashnikov, S. Mehrotra, P. Smyth, and N. Venkatasubramanian, Situational technologies for disaster response, in *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*, Chen et al (eds), Springer, 2007, in press.

BC2 E. Ip, I. Cadez, and P. Smyth, 'Psychometric methods of latent variable modeling', in *Handbook of Data Mining*, N. Ye (ed.), Erlbaum Associates, 215-246, 2003.

BC3 P. Smyth, 'Data mining at the interface of computer science and statistics,' invited chapter for *Mining Scientific Data Sets*, Grossman, R., Kamath, C., Kumar, V., and Nambur, R. (eds.), Kluwer Academic, 35-61, 2001.

- BC4 P. Smyth, 'Hidden Markov models,' in *The MIT Encyclopaedia of the Cognitive Sciences*, R. A. Wilson and F. C. Keil (eds.), Cambridge, MA: The MIT Press, 373–374, 1999, invited contribution. (This book was awarded "best psychology title published in 1999" by the American Association of Publishers).
- BC5 W. R. Shankle, S. Mani, M. Pazzani, and P. Smyth, 'Dementia screening with machine learning methods,' in *Intelligent Data Analysis in Medicine and Pharmacology*, Elpida Keravnou, Nada Lavrac and Blaz Zupan (eds.), Kluwer Academic Publishers, 1998.
- BC6 U. M. Fayyad, P. Smyth, M. C. Burl, and P. Perona, 'A learning approach to object recognition: applications in science image database exploration and analysis,' in *Early Visual Learning*, S. Nayar and T. Poggio (eds.), pp.237–268, 1996.
- BC7 P. Smyth, M. Burl, U. M. Fayyad, P. Perona, 'Knowledge discovery in large image databases: dealing with uncertainties in ground truth,' in *Advances in Knowledge Discovery and Data Mining*, U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurasamy (eds.), AAAI/MIT Press, pp.517–539, 1996.
- BC8 U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, 'From data mining to knowledge discovery: an overview,' in *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurasamy (eds.), Palo Alto, CA: AAAI/MIT Press, pp.1–34, 1996.
- BC9 P. Smyth, 'Learning with probabilistic supervision,' in *Computational Learning Theory and Natural Learning Systems 3*, T. Petsche, S. Hanson, and J. Shavlik (eds), Cambridge, MA: MIT Press, pp.163–182, 1995.
- BC10 U. M. Fayyad and P. Smyth, 'The automated analysis, cataloguing, and searching of digital image libraries: a machine learning approach,' in *Advances in Digital Libraries*, N. R. Adam and B. Bhargava (eds.), *Lectures Notes in Computer Science*, Springer-Verlag, pp.225-249, 1995.
- BC11 P. Smyth, 'Detecting novel fault conditions with hidden Markov models and neural networks,' in *Pattern Recognition in Practice IV: Multiple Paradigms, Comparative Studies, and Hybrid Systems*, E. S. Gelsema and L. N. Kanal (eds.), Elsevier : Amsterdam, pp.525–536, 1994.
- BC12 P. Smyth, 'Probability density estimation and local basis function neural networks,' in *Computational Learning Theory and Natural Learning Systems 2*, S. Hanson, T. Petsche, M. Kearns, R. Rivest (eds), Cambridge, MA: MIT Press, pp.233–248, 1994.
- BC13 P. Smyth, 'Admissible stochastic complexity models for classification problems,' in *Artificial Intelligence Frontiers in Statistics: AI and Statistics 3*, D. Hand (ed.), Chapman & Hall: London, pp.335–347, 1993 (same paper as number J14 under journal publications list).
- BC14 P. Smyth and R. M. Goodman, 'Rule induction using information theory,' in *Knowledge Discovery in Databases*, G. Piatetsky-Shapiro and W. Frawley (eds.), The MIT Press, Cambridge: MA, pp. 159–176, 1991.
- BC15 P. Smyth, J. Statman, G. Oliver and R. Goodman, 'Combining knowledge-based techniques and simulation with applications to communications network management,' in *Integrated Network Management II*, I. Krishnan and W. H. Zimmer (eds.), Elsevier Science Publishers, April 1991.

Conference Papers, peer-reviewed

- P1 A. Asuncion, P. Smyth, M. Welling, and Y. Teh 'On the role of smoothing in topic models,' *Proceedings of the 25th Conference on Uncertainty in AI*, June 2009.
- P2 A. Asuncion, P. Smyth, and M. Welling, 'Asynchronous distributed learning of topic models,' *Neural Information Processing Systems (NIPS) 21*, MIT Press, pp.81–88, December 2008.
- P3 C. Chemudugunta, P. Smyth, and M. Steyvers, 'Combining concept hierarchies and statistical topic models,' *Proceedings of the Conference on Information and Knowledge Management (CIKM-08)*, October 2008.
- P4 C. Chemudugunta, A. Holloway, P. Smyth, and M. Steyvers, 'Modeling documents by combining semantic concepts with unsupervised statistical learning,' in *Proceedings of the International Semantic Web Conference (ISWC-08)*, Springer Verlag, Berlin, pp. 229–244, October 2008.
- P5 J. Hutchins, A. Ihler, and P. Smyth, 'Probabilistic analysis of a large-scale urban traffic data set,' *Second International Workshop on Knowledge Discovery from Sensor Data (ACM SIGKDD Conference, KDD-08)*, August 2008.
- P6 I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, M. Welling, 'Fast collapsed Gibbs sampling for latent Dirichlet allocation,' *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp. 569–577, August 2008.
- P7 J. Hutchins, A. Ihler, P. Smyth, 'Modeling count data from multiple sensors: a building occupancy model,' in *Computational Advances in Multisensor Adaptive Processing (CAMSAP)*, IEEE Press, pp.241–244, 2007.
- P8 D. Newman, A. Asuncion, P. Smyth, M. Welling, 'Distributed inference for latent Dirichlet allocation', *Advances in Neural Information Processing Systems 20*, MIT Press, pp. 1081–1088, December 2007.
- P9 S. Kirshner and P. Smyth, 'Infinite mixtures of trees,' in *Proceedings of the 24th International Conference on Machine Learning*, ACM International Conference Proceeding Series, New York, pp.417–723, June 2007.
- P10 D. Newman, K. Hagedorn, C. Chemudugunta, and P. Smyth, 'Subject metadata enrichment using statistical topic models,' in *Joint Conference on Digital Libraries*, June 2007.
- P11 A. Ihler and P. Smyth, Learning time-intensity profiles of human activity using non-parametric Bayesian models, *Advances in Neural Information Processing Systems 19*, December 2006.
- P12 C. Chemudugunta, P. Smyth, and M. Steyvers, Modeling general and specific aspects of documents with a probabilistic topic model, *Advances in Neural Information Processing Systems 19*, MIT Press, pp. 241–248, 2006.
- P13 S. Kim and P. Smyth, Hierarchical Dirichlet processes with random effects, *Advances in Neural Information Processing Systems 19*, December 2006.
- P14 S. Kim, P. Smyth, and H. Stern, A nonparametric Bayesian approach to detecting spatial activation patterns in fMRI data, *Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, Berlin: Springer-Verlag, 217–224, October 2006.

- P15 D. Newman, C. Chemudugunta, P. Smyth, and M. Steyvers, Statistical entity-topic models, *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, 680–686, August 2006.
- P16 A. Ihler, J. Hutchins, and P. Smyth, Adaptive event detection with time-varying Poisson processes, *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, 207–216, August 2006.
- P17 I. Porteous, A. Ihler, P. Smyth, M. Welling, Gibbs sampling for (coupled) infinite mixture models in the stick-breaking representation, in *Proceedings of the Uncertainty in AI Conference*, July 2006.
- P18 D. Newman, C. Chemudugunta, P. Smyth, and M. Steyvers, Analyzing entities and topics in news articles using statistical topic models, *IEEE International Conference on Intelligence and Security Informatics*, Springer Lecture Notes in Computer Science (LNCS) 3975, 93–104, May 2006.
- P19 S. Kim, P. Smyth, H. Stern, J. Turner, Parametric response surface models for analysis of multi-site fMRI data, in *Proceedings of 8th International Conference on Medical Image Computing and Computer Assisted Intervention*, Springer Lecture Notes in Computer Science 3749, 352–359, October 2005.
- P20 S. White and P. Smyth, A spectral clustering approach to finding communities in graphs, in *Proceedings of the SIAM International Conference on Data Mining*, Newport Beach, CA, SIAM Press, April 2005.
- P21 S. Gaffney and P. Smyth, 'Joint probabilistic curve-clustering and alignment,' *Advances in Neural Information Processing 17 (Proceedings of the 2004 Conference)*, MIT Press, 473–480, 2005.
- P22 M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, 'Probabilistic author-topic models for information discovery,' in *Proceedings of the Tenth ACM International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, 306–315 August 2004.
- P23 M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, 'The author-topic model for authors and documents,' in *Proceedings of the 20th International Conference on Uncertainty in AI*, ACM International Conference Proceeding Series, 487–494, 2004.
- P24 S. Kim, P. Smyth, and S. Luther, 'Modeling waveform shapes with random effects segmental hidden Markov models,' in *Proceedings of the 20th International Conference on Uncertainty in AI*, ACM International Conference Proceeding Series, 309–316, 2004.
- P25 S. Kirshner, P. Smyth, and A. Robertson, 'Conditional Chow-Liu tree structures for modeling discrete-valued vector time series,' in *Proceedings of the 20th International Conference on Uncertainty in AI*, ACM International Conference Proceeding Series, 317–324, 2004.
- P26 S. J. Camargo, A. W. Robertson, S. J. Gaffney, and P. Smyth, 'Cluster analysis of the Western North Pacific tropical cyclone tracks,' *Proceedings of the 26th Conference on Hurricanes and Tropical Meteorology*, 3–7 May 2004, Miami, FL, 10A.7, pp. 250–251.
- P27 D. Chudova, C. Hart, E. Mjolsness and P. Smyth, 'Gene expression clustering with functional mixture models,' *Neural Information Processing Conference (NIPS 2003)*, Vancouver, December 2003; also in *Advances in Neural Information Processing 16*, MIT Press, 2004.

- P28 S. White and P. Smyth, 'Algorithms for estimating relative importance in networks,' in *Proceedings of the Ninth ACM International Conference on Knowledge Discovery and Data Mining*, Washington DC, pp. 266–275. August 2003.
- P29 D. Chudova, S. Gaffney, E. Mjolsness and P. Smyth, 'Translation-invariant mixture models for curve-clustering,' in *Proceedings of the Ninth ACM International Conference on Knowledge Discovery and Data Mining*, Washington DC, pp. 79–88, August 2003.
- P30 S. Kirshner, S. Parise and P. Smyth, 'Unsupervised learning from permuted data,' in *Proceedings of the Twentieth International Conference on Machine Learning, ICML-03*, Washington DC, pp. 345–352, August 2003.
- P31 D. Chudova, S. Gaffney and P. Smyth, 'Probabilistic models for joint clustering and time-warping of multidimensional curves,' in *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers, 134–141, August 2003.
- P32 D. Pavlov and P. Smyth, 'Approximate query answering by model averaging,' SIAM International Conference on Data Mining, April 2003.
- P33 S. Gaffney and P. Smyth, 'Curve clustering with random effects mixtures,' in C. M. Bishop and B. J. Frey (eds), *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Jan 3-6, 2003, Key West, FL.
- P34 X. Ge, S. Parise, and P. Smyth, 'Clustering Markov states into equivalence classes using SVD and heuristic search algorithms,' in C. M. Bishop and B. J. Frey (eds), *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Jan 3-6, 2003, Key West, FL.
- P35 S. Kirshner, I. Cadez, P. Smyth, and C. Kamath, 'Learning to classify galaxy shapes using EM,' Neural Information Processing Conference (NIPS 2002), Vancouver, December 2002: MIT Press.
- P36 S. Kirshner, I. Cadez, P. Smyth, E. Cantu-Paz, and C. Kamath, 'Probabilistic model-based detection of bent-double radio galaxies,' *Proceedings of the International Conference on Pattern Recognition*, August 2002.
- P37 D. Chudova and P. Smyth, 'Pattern discovery in sequences under a Markov assumption,' in *Proceedings of the ACM Eighth International Conference on Knowledge Discovery and Data Mining*, July 2002 (winner, best research paper award).
- P38 S. Scott and P. Smyth, 'The Markovian Poisson cascade with applications to web traffic modeling,' *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (eds.), Oxford University Press, 2003.
- P39 I. Cadez and P. Smyth, 'Bayesian predictive profiles with applications to retail transaction data,' Neural Information Processing Conference (NIPS 2001), Vancouver, December 2001: MIT Press.
- P40 I. Cadez, P. Smyth, and H. Mannila, 'Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction,' in *Proceedings of the ACM Seventh International Conference on Knowledge Discovery and Data Mining*, ACM: New York, NY, pp. 37–46, August 2001.
- P41 D. Pavlov and P. Smyth, 'Probabilistic query models for transaction data,' in *Proceedings of the ACM Seventh International Conference on Knowledge Discovery and Data Mining*, ACM: New York, NY, pp. 164–173, August 2001.

- P42 I. Cadez and P. Smyth, 'Model complexity, goodness-of-fit, and diminishing returns,' presented at the Neural Information Processing Conference (NIPS 2000), Denver, CO, November 2000: MIT Press, pp. 388–394.
- P43 X. Ge and P. Smyth, 'Deformable Markov model templates for time-series pattern-matching,' in *Proceedings of the ACM Sixth International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM Press, pp.81–90, August 2000 (runner-up, best research paper award).
- P44 I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White, 'Visualization of navigation patterns on a Web site using model-based clustering,' in *Proceedings of the ACM Sixth International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM Press, pp. 280–284, August 2000.
- P45 I. Cadez, S. Gaffney, and P. Smyth, 'A general probabilistic framework for clustering individuals,' in *Proceedings of the ACM Sixth International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM Press, pp. 140–149, August 2000.
- P46 D. Pavlov, D. Chudova, and P. Smyth, 'Towards scalable support-vector machines using squashing,' in *Proceedings of the ACM Sixth International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM Press, pp. 295–299, August 2000.
- P47 D. Pavlov, H. Mannila, and P. Smyth, 'Probabilistic models for query approximation with large sparse binary data sets,' in *Proceedings of the 2000 Uncertainty in AI Conference*, San Francisco, CA: Morgan Kaufmann, pp. 465–472, July 2000.
- P48 H. Mannila and P. Smyth, 'Approximate query answering with frequent sets and maximum entropy,' *Proceedings of ICDE 2000*, IEEE Press, 309, February 2000.
- P49 S. Gaffney and P. Smyth, 'Trajectory clustering using mixtures of regression models,' in *Proceedings of the ACM 1999 Conference on Knowledge Discovery and Data Mining*, S. Chaudhuri and D. Madigan (eds.), New York, NY: ACM, 63–72, August 1999.
- P50 H. Mannila, D. Pavlov, and P. Smyth, 'Prediction with local patterns using cross-entropy,' in *Proceedings of the ACM 1999 Conference on Knowledge Discovery and Data Mining*, S. Chaudhuri and D. Madigan (eds.), New York, NY: ACM, 357–361, August 1999.
- P51 X. Ge, W. Pratt, and P. Smyth, 'Discovering Chinese words from unsegmented text,' in *Proceedings of 22nd International Conference on Research and Development in Information Retrieval (SIGIR '99)*, M. Hearst, F. Gey, R. Tong (eds.), New York, NY: ACM, 271–272, August 1999.
- P52 I. V. Cadez, C. E. McLaren, P. Smyth, and G. J. McLachlan 'Hierarchical models for screening of iron-deficient anemia,' in *Proceedings of the 1999 International Conference on Machine Learning*, I. Bratko and S. Dzeroski (eds.), Los Gatos: CA, Morgan Kaufmann, 77–86, June 1999.
- P53 P. Smyth, 'Probabilistic model-based clustering of multivariate and sequential Data,' in *Proceedings of the Seventh International Workshop on AI and Statistics*, D. Heckerman and J. Whittaker (eds.), Los Gatos, CA: Morgan Kaufmann, 299–304, January 1999.
- P54 G. Das, K. Lin, H. Mannila, G. Rengenathan, and P. Smyth, 'Rule discovery from time series,' *Proceedings of the 1998 Conference on Knowledge Discovery and Data Mining*, R. Agrawal and P. Stolorz (eds.), Menlo Park, CA: AAAI Press, 16–22, 1998 (runner-up, best research paper award).

- P55 P. Smyth and D. Wolpert, 'Stacked density estimation,' accepted for oral presentation, *Neural Information Processing System Conference*, Denver, CO, November 1997: also in *Advances in Neural Information Processing 10*, April 1998.
- P56 P. Smyth, M. Ghil, K. Ide, J. Roden, and A Fraser, 'Detecting atmospheric regimes using cross-validated clustering,' *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Menlo Park, CA: AAAI Press, 61–66, 1997 winner, best applied research paper award.
- P57 P. Smyth and D. Wolpert, 'Anytime exploratory data analysis for massive data sets,' *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Menlo Park, CA: AAAI Press, 54–60, 1997.
- P58 E. Keogh and P. Smyth 'A probabilistic approach to fast pattern matching in time series databases,' *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Menlo Park, CA: AAAI Press, 24–30, 1997.
- P59 W. R. Shankle, Mani, S., Pazzani, M. J. and Smyth, P., 'Use of a computerized patient record database of normal aging and very mildly demented subjects to compare classification accuracies obtained with machine learning methods and logistic regression,' *Computing Science and Statistics*, 29(2), 201–209, 1997.
- P60 W. R. Shankle, S. Mani, M. Pazzani, and P. Smyth, 'Detecting very early stages of dementia using machine learning methods,' in *Lecture Notes in Artificial Intelligence: Artificial Intelligence in Medicine, AIME97*, Springer, pp.73–85, 1997.
- P61 P. Smyth, 'Cross-validated likelihood for model selection in unsupervised learning,' in *Proceedings of the Sixth International Workshop on AI and Statistics*, 473–480, January 1997.
- P62 P. Smyth, 'Clustering sequences using hidden Markov models,' in *Advances in Neural Information Processing 9*, M. C. Mozer, M. I. Jordan and T. Petsche (eds.), Cambridge, MA: MIT Press, 648–654, 1997.
- P63 U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, 'Knowledge discovery and data mining: towards a unifying framework' in *Proceedings of the 1996 Knowledge Discovery and Data Mining Conference*, AAAI Press, 82–88, August 1996.
- P64 P. Smyth, 'Clustering using Monte-Carlo cross-validation,' in *Proceedings of the 1996 Knowledge Discovery and Data Mining Conference*, Menlo Park: CA, AAAI Press, 126–133, 1996.
- P65 P. Smyth, A. Gray, and U. M. Fayyad, 'Retrofitting decision tree classifiers using kernel density estimation,' in *Proceedings of the 1995 Conference on Machine Learning*, Morgan Kaufman, 506–514, 1995.
- P66 P. Smyth, M.C. Burl, U. M. Fayyad, P. Perona, P. Baldi, 'Inferring ground truth from subjectively-labeled images of Venus,' in *Advances in Neural Information Processing Systems 7*, G. Tesauro, D. S. Touretzky, and T. K Leen (eds.), MIT Press, 1085–1092, 1995: presented at the 1994 Neural Information Processing Conference, Denver, CO, December 1994.
- P67 M.C. Burl, U. M. Fayyad, P. Perona, P. Smyth, 'Automated analysis of radar imagery of Venus: handling lack of ground truth,' in *Proceedings of the IEEE International Conference on Image Processing*, vol.III, pp.236–240, 1994.

- P68 K. M. Cheung and P. Smyth, 'Adaptive source-coding for geometrically distributed integer alphabets,' in *Proceedings of the IEEE Data Compression Conference*, Snowbird, Utah, April 1994.
- P69 M.C. Burl, U. M. Fayyad, P. Perona, P. Smyth, M. P. Burl, 'Automating the hunt for volcanoes on Venus,' *Proceedings of the 1994 Computer Vision and Pattern Recognition Conference (CVPR-94)*, Los Alamitos, CA: IEEE Computer Society Press, pp.302–309, 1994.
- P70 Z. Zheng, R. Goodman, and P. Smyth, 'Discrete recurrent neural networks as pushdown automata,' in *Proceedings of the International Symposium on Nonlinear Theory and its Applications*, Hawaii, vol.3, pp.1033-1038, December 1993.
- P71 P. Smyth, 'Probabilistic anomaly detection in dynamic systems,' presented at the 1993 Neural Information Processing Conference, Denver, CO, December 1993: also in *Advances in Neural Information Processing Systems 6*, J. D. Cowan, G. Tesauro, J. Alspector (eds.), Morgan Kaufmann Publishers:San Mateo, pp.825–832., 1994.
- P72 P. Smyth, 'Hidden Markov models and neural networks for fault detection in dynamic systems,' in *Neural Networks for Signal Processing III*, IEEE Press: New York, pp. 582–591, 1993: presented at the 1993 IEEE Workshop on Neural Networks and Signal Processing, Baltimore, September 1993.
- P73 Z. Zheng, R. Goodman, and P. Smyth, 'Self-clustering recurrent networks,' in *Proceedings of the IEEE International Joint Conference on Neural Networks*, San Francisco, March 1993.
- P74 P. Smyth and J. Mellstrom, 'Detecting novel classes with applications to fault diagnosis,' in *Proceedings of the Ninth International Conference on Machine Learning*, Morgan Kaufmann Publishers: Los Altos, CA, 1992, pp.416–425.
- P75 P. Smyth and J. Mellstrom, 'Fault diagnosis of antenna pointing systems using hybrid neural networks and signal processing techniques,' presented at the IEEE Neural Information Processing Systems Conference, Denver, CO, December 1991: also in *Advances in Neural Information Processing Systems 4*, R. Lippmann (ed.), Morgan Kaufmann Publishers: Los Altos, CA, 667–674, 1992.
- P76 R. M. Goodman, J. W. Miller, and P. Smyth, 'Objective functions for probability estimation,' in *Proceedings of the 1991 International Joint Conference on Neural Networks*, Seattle, July 1991, vol.1, pp.881–886.
- P77 P. Smyth, 'On admissible stochastic complexity models for neural network classifiers,' Neural Information Processing Conference, Denver, CO, December 1990: also in *Advances in Neural Information Processing Systems 3*, R. Lippmann, J.E.Moody, D.S. Touretzky, (eds.), Morgan Kaufmann Publishers:San Mateo, CA, pp.818–824, 1991.
- P78 P. Smyth, R. M. Goodman, and C. Higgins, 'A hybrid rule-based/Bayesian classifier,' in *Proceedings of the Ninth European Conference on Artificial Intelligence*, Pitman Publishing: London, pp 610–615, 1990.
- P79 R. M. Goodman, J. W. Miller and P. Smyth, 'An information-theoretic approach to rule-based connectionist expert systems,' Neural Information Processing Systems Conference, Denver, CO, December 1989: also in *Advances in Neural Information Processing Systems 1*, D. Touretzky, (ed.), Morgan Kaufmann Publishers: Los Altos, CA, 256–263, 1989.

P80 R. M. Goodman and P. Smyth, 'Information-theoretic rule induction,' *Proceedings of the 1988 European Conference on Artificial Intelligence*, Pitman: London, 1988.

P81 E. C. Posner and P. Smyth, 'Test access in multi-stage switching networks,' *Proceedings of the 12th International Teletraffic Congress*, Turin, Italy, June 2-10th, 1988: also in *Teletraffic Science for New Cost-Effective Systems, Networks, and Services*, M. Bonatti (ed.), North-Holland Studies in Telecommunications, vol. 12, Amsterdam: North Holland, 1989.

Patents

U. S. Patent no. 5465321, *Hidden Markov Models for Fault Detection in Dynamic Systems*, assigned to NASA, inventor is P. Smyth, issued November 7 1995.

U. S. Patent no. 48707280, *Cross-connect switch*, assigned to Pacific Bell, inventors are E. C. Posner and P. Smyth, issued February 21 1989.