PADHRAIC SMYTH

Department of Computer Science, Bren Hall 4216
School of Information and Computer Sciences
University of California, Irvine
CA 92697-3435
telephone: (949) 824 2558
email: smyth@ics.uci.edu

Professional Positions

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

- Distinguished Professor: 2023 to present
- Chancellor's Professor: 2018 to 2023
- Full Professor: July 2003 to 2018
- Associate Professor: July 1998 to June 2003
- Assistant Professor: April 1996 to June 1998

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.

Additional Professional Roles and Affiliations

Joint Faculty Appointments:

- Department of Statistics, UC Irvine, July 2008–present.
- Department of Education, UC Irvine, July 2017–present.

Founding Director, UCI Data Science Initiative, University of California, Irvine, July 2014–June 2018.

Founding Director, Center for Machine Learning and Intelligent Systems, University of California, Irvine, January 2007–June 2014.

Faculty Member, Institute for Genomics and Bioinformatics (IGB), UC Irvine, Member 2001–present.

Faculty Member, Institute for Mathematical Behavioral Sciences (IMBS), UC Irvine, 1999–2022.

Faculty Member, Center for Digital Transformation, UC Irvine, 2012–present.

Faculty Member, Program for Mathematical, Computational, and Systems Biology (MCB), UC Irvine, 2007–present.

Faculty Member, Center for Research on Information Technology and Organizations (CRITO), UC Irvine, 2008–2012.

Founding Director and Executive Committee 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.

Honors and Awards

Fellow, American Association for the Advancement of Science (AAAS), elected 2022

Fellow, Association for Computing Machinery (ACM), elected 2013

Fellow, Association for the Advancement of Artificial Intelligence (AAAI), elected 2010

ACM SIGKDD Innovation Award, 2009

Best paper awards: ACM SIGKDD Conference (best paper (1997, 2002), runner-up best paper (1998, 2000)), ACM/IEEE Joint Conference on Digital Libraries (JCDL) (shortlist for best paper, 2007), Educational Data Mining Conference (best paper, 2018)

Qualcomm Faculty Awards, 2019/2020/2021

Google Faculty Research Awards, 2008 and 2014

IBM Faculty Partnership Award, 2001

National Science Foundation CAREER award, 1997

ACM Teaching Award, UC Irvine, 1997

NASA Group Achievement award, Jet Propulsion Labaratory, 1997

Lew Allen Award for Excellence in Research, Jet Propulsion Laboratory, 1993

17 NASA Certificates for Technical Innovation (1991–1996)

Advisory and Consulting Activities

Smith Baluch LLP (2022-present); Cove Fund (2021-present); Candor Technologies (2021-present); Fox Rothschild LLP (2021); Fish and Richardson (2021); AdvanceOC Advisory Board (2020-present); Wilson, Sonsoni, Goodrich and Rosati (2019-2021); Fenwick and West LLP (2019-2021); QuinnEmanuel LLP (2019-2020); Morgan Lewis and Bockius LLP (2019); Erise IP (2017-2018; Toshiba (2018-2019), First American (2018-2019); ProLung, Inc (2017-2019); Unified Patents (2016-2019); University of Washington (2016-2019); Klarquist LLP (2015-2016); Frost Data Capital (2014-2015); AST Inc (2013-2015); Samsung (2012-2015); SOCCCD (2012-present); DigitalRisk (2010-2012); CoreLogic (2011-2014); IdentityMetrics (2010-2012); Microsoft (2010-2011); ImageCat (2010); eBay (2009-2011); DataAnalytics LLC (2009-2011); QuinnEmanuel LLP (2011); Latham and Watkins (2008-2009, 2011); Netflix (2006-2009); 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 (1998); AT&T (1996-1998).

Postdoctoral Advisees and Current Positions

Ralf Krestel, 2011-2013; Professor, Kiel University, Germany.
Tracy Holsclaw, 2011-2014; Consultant, San Jose, CA.
Romain Thibaux, 2008-2009; Software Engineer, Waymo, Mountain View, CA.
Alex Ihler, 2005-2006; Professor, Department of Computer Science, UC Irvine.
Michael Duff, 2005-2006; Researcher, Fred Hutchinson Cancer Research Center, Seattle, WA.
Michal Rosen-Zvi, 2003-2004; Director, AI for Healthcare and Lifescience, IBM Research, Israel.

PhD Students

PhD Advisees and Current Positions

Robert Logan IV (co-advised with Sameer Singh), PhD 2022; Dataminr, New York
Disi Ji, PhD 2020; Instagram, Menlo Park, CA
Chris Galbraith, PhD 2020; Mandiant, Philadelphia, PA
Jihyun Park, PhD 2019; Apple, Cupertino, CA
Dimitris Kotzias, PhD 2018; Google, Zurich
Eric Nalisnick, PhD 2018; Assistant Professor, University of Amsterdam
Moshe Lichman, PhD 2017; Google, Irvine, CA
Nick Navaroli, PhD 2014; Google, Irvine, CA
Jimmy Foulds, PhD 2014: Assistant Professor, Department of Computer Science, UMBC
Chris DuBois, PhD 2013: Apple, Seattle
America Chambers, PhD 2013: Assistant Professor, Department of Mathematics and Computer Science, University of Puget Sound
Drew Frank (co-advised with Alex Ihler), PhD 2013: Apple, Seattle
Arthur Asuncion, PhD 2011: Google, Seattle, WA
Jon Hutchins (co-advised with Alex Ihler), PhD 2010: Google, Pittsburgh, PA
Chaitanya Chemudugunta, PhD 2009: Director, Data Science/Research, Pandora, CA
Seyoung Kim, PhD 2007: Associate Professor, Department of Bioinformatics, CMU, Pittsburgh
Darya Chudova, PhD 2007: Senior VP of Technology, Guardant Health, Redwood City, CA
Sergey Kirshner, PhD 2005: Amazon, Palo Alto, CA
Scott Gaffney, PhD 2004: VP of Search Engineering, eBay, San Jose, CA
Xianping Ge, PhD 2002
Igor V. Cadez, PhD 2002
Dimitry Pavlov, Consultant, PhD 2001

Current PhD Students

Advanced to Candidacy: Alex Boyd (co-advised with Stephan Mandt), Gavin Kerrigan, Rachel Longjohn, Markelle Kelly
Pre-Candidacy: Sam Showalter, Catarina Belem (co-advised with Sameer Singh), Yuxin Chang, Giosue Migliorini

Professional Activities

Journals: Associate/Action Editor

ACM Transactions on Knowledge Discovery and Data, guest editor of special issue on best papers from *ACM SIGKDD 2011 Conference*, TKDD 6(4), 2012.

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, Book Series, Centers: Advisory/Editorial Board Member

Journal of Machine Learning Research, 2000-2020.

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

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

Bayesian Analysis, 2004-2007.

Insight Center for Data Analytics, University College Dublin, Scientific Advisory Member, 2015-2020.

Conference Program and General Chair Positions

Associate Program Chair, International Joint Conference on Artificial Intelligence (IJCAI), 2022

Program Chair for the Uncertainty in Artificial Intelligence (UAI) Conference, 2013.

Program Chair for 17th ACM SIGKDD Conference, San Diego, 2011.

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

General Chair for the Sixth International Conference on Artificial Intelligence and Statistics, January 1997.

Other Conference and Workshop Organization Roles

Conference Organization Roles: Senior Area Chair/Area Chair, NeurIPS 2017, 2018, 2019, 2020,2021; Senior Area Chair/Area Chair, ICML 2018, 2019, 2020, 2021,2022,2023; Senior Area Chair, AAAI 2020; Panels Chair for ACM SIGKDD Fifth International Conference on Knowledge Discovery and Data Mining, 1999; Tutorials co-Chair for National Conference on Artificial Intelligence, 1998; Tutorials Chair for the ACM SIGKDD Conferences on Knowledge Discovery and Data Mining, 1997 and 1998; Publicity Chair for the ACM SIGKDD Conferences on Knowledge Discovery and Data Mining, 1995 and 1996.

Workshop Co-Chair/Organizer for: Dagstuhl Seminar, Automating Data Science, 2018; Workshop on Algorithmic and Statistical Approaches for Large Social Network Data Sets, NIPS Conference, Lake Tahoe, 2012; Workshop on User-Centered Modeling, Institute for Mathematics and its Applications (IMA), University of Minnesota, 2012.; Workshop on Scientific Data Mining, Institute for Pure and Applied Mathematics (IPAM), UCLA, 2002; Workshop on Temporal and Spatial Machine Learning, International Conference on Machine Learning (ICML), 2001; Massive Datasets workshop at the 1998 Neural Information Processing Conference (NIPS).

Research and Training Grants, Contracts and Gifts

81. *AI/ML and Data Science Training Datasets*, subaward to NIH 3OT2OD032581-01S1, $295,021, Jan 1 2023 to Sept 16 2023, Principal Investigator.

80. *Improving Prediction of Fire Extremes in the GEOS Forecasting System on Daily and Seasonal Timescales*, NASA, Sept 1 2021 to June 30 2025, $1,040,166, Co-principal investigator (PI: Jim Randerson, Earth System Sciences, UCI).

79. *Fair Risk Predictions for Underrepresented Populations using Electronic Health Records*, NIH R01AG065330-02S1, Sept 1 2021 to April 30 2022, $167,792, co-investigator, (PI: Judy Zhong, Biostatistics, NYU).

78. *Data Science Training and Practices: Preparing a Diverse Workforce via Academic and Industrial Partnership*, NSF IIS-2123366, Sept 1 2021 to Aug 31 2024, $751,921, Co-principal investigator (PI: Babak Shahbaba, Statistics, UCI).

77. *Personalized Risk Predictions with Deep Learning Methods in the Presence of Missing and Biased Electronic Health Record Data*, NIH R01-LM013344, Aug 6 2021 to May 31 2025, $498,957 (UCI portion), Principal Investigator (MPI with Judy Zhong, Biostatistics, NYU).

76. *Center of Excellence in Forensic Statistics (CSAFE2)*, National Institute of Standards and Technology (NIST), award number 70NANB20H019 , $20,000,000 ($4,000,000 for UC Irvine), June 2020 to May 2025; co-Investigator (UCI PI: Hal Stern).

75. *HPI Research Center in Machine Learning and Data Science at UC Irvine*, Hasso Plattner Institute (gift), April 2020 to Dec 2024, $3,592,500, Co-principal investigator (PI: Erik Sudderth, UCI).

74. *Addressing the Critical Role of Innate/Adaptive Immunity by Integrating Novel Informatics, Translation Technologies and Ongoing Clinical Trial Research*, NIH 3UL1TR001414-06S1, Sept 2020 to June 2021, $ 1,088,735, co-investigator (PI: Dan Cooper, School of Medicine, UCI).

73. *Analyzing Information Exchange in Human-Human Dialog using Machine Learning*, SAP Innovation Center, $124,000, April 1 2020 to March 31 2021, Principal Investigator.

72. *Generative Expectation-based Response and Novelty Identification*, DARPA/SRI-HR001120C0021, $1,087,251, Oct 1 2019 to May 30 2022, Co-investigator (PI: Stephan Mandt, Computer Science, UCI).

71. *Machine Learning Democratization via a Linked, Annotated, Repository of Datasets*, National Science Foundation (CCRI: ENS), award number NSF-1925741, $1,792,952, Oct 1 2019 to Sept 30 2022. Co-principal investigator (PI: Sameer Singh, Computer Science, UCI).

70. *Hybrid Human Algorithm Predictions: Balancing Effort, Accuracy, and Perceived Autonomy*, National Science Foundation (EAGER: AI-DCL), award number NSF-1927245, $293,923, Aug 15 2019 to Aug 14 2021. Co-principal investigator (PI: Mark Steyvers, Cognitive Sciences, UCI).

69. *Assessment of Machine Learning Algorithms in the Wild*, National Science Foundation, award number NSF-1900644, $1,199,898, Oct 1 2019 to Sept 30 2023, Principal Investigator.

68. *Qualcomm Faculty Award*, $225,000 (gift), May 2019/March 2022, Principal Investigator.

67. *Innovation Center for Advancing Ecosystem Climate Solutions*, California Strategic Growth Council, award number CCR20021, $4,604,140, 4/01/2019 to 3/31/2022, co-investigator (PI: Mike Goulden, Earth Systems Sciences, UCI).

66. *Hands-free Documentation in Clinical Practice*, SAP, $172,000 (gift/sponsored project), October 2018, co-Principal Investigator (with Kai Zheng, Department of Informatics, UCI).

65. *TRIPODS-X: Data Science Frontiers in Climate Science*, National Science Foundation, award number NSF-1839336, $300,000, Oct 1 2018 to Sept 30 2021, co-PI (PI: Efi Foufoula-Georgiou, Civil and Environmental Engineering, UCI).

64. *Large-Scale Classification Algorithms*, eBay Labs, $30,000 (gift), Dec 1 2017, Principal Investigator.

63. *Center for Machine Learning and Intelligent Systems*, Cylance, $50,000 (gift), Dec 1 2017, Principal Investigator.

62. *Development of Computational Methods for Evaluating Patient-Doctor Communication*, PCORI, $270,000 (UCI portion), award number ME-1602-34167, July 1 2017 to June 30th 2019, co-Investigator (PI: Zac Imel, U Utah).

61. *NRT-DESE: Team Science for Integrative Graduate Training in Data Science and Physical Science*, NSF, award number NSF-1633631, Sep 15 2016 to Aug 31 2021, $2,967,150, Principal Investigator.

60. *Learning Individual Predictive Choice Models*, Adobe Research Award, $50,000, October 2016, Principal Investigator.

59. *Transformative Computational Infrastructures for Cell-Based Biomarker Diagnostics*, NIH, award number U01TR001801-01, 09/01/16 08/31/21, $766,000 (UCI portion), co-Investigator (PI: Richard Scheuermann, Venter Institute/UCSD).

58. *The Big DIPA: Data Image Processing and Analysis*, NIH BD2K Program, award number 1R25EB022366-01, $486,000, Sept 30 2015 to June 30th 2018, co-Investigator (UCI PI: Charless Fowlkes).

57. *Investigating Virtual Learning Environments*, National Science Foundation, award number NSF-1535300, $2,500,000, Oct 1 2015 to Sept 30th 2020, co-Investigator (UCI PI: Mark Warschauer).

56. *Center of Excellence in Forensic Statistics (CSAFE2)*, National Institute of Standards and Technology (NIST), award number 70NANB15H176, $20,000,000 ($3,700,000 for UC Irvine), June 2015 to May 2020; co-Investigator (UCI PI: Hal Stern).

55. *Data-Intensive Research and Education Center in Science, Technology, Engineering, and Mathematics (DIRECT-STEM)*, NASA MIRO program, award number NNX15AQ06A, $5,000,000 ($1,250,000 for UC Irvine), Sept 1 2015 to Aug 31st 2020, Principal Investigator.

54. *Analyzing Individual Event Data over Time*, Google Faculty Research Award, $60,000, March 2014, Principal Investigator.

53. *Peer Assessment and Academic Achievement in a Gateway MOOC*, Bill and Melinda Gates Foundation, Oct 1 2013, $25,000, Co-Investigator (PI: Mark Warschauer, UC Irvine).

52. *Statistical Learning Algorithms for Micro-Event Time Series Data*, National Science Foundation, award number IIS-1320527, Oct 1 2013 to Sept 30th 2018, $499,880, Principal Investigator.

51. *Balancing the Portfolio: Efficiency and Productivity of Federal Biomedical R&D Funding*, National Science Foundation, award number 1158699, Aug 15 2012 to July 31 2015, $297,331, Principal Investigator (original PI, David Newman).

50. *Location-based Social Media for Context-based Analysis of Transportation Data*, Xerox UAC Research Award, Jan 1st 2013 to Dec 31st 2015, $90,000 gift, Principal Investigator.

49. *Collaborative Research, Type 1: Decadal Prediction and Stochastic Simulation of Hydroclimate over Monsoonal Asia*, US Department of Energy, award number DOE SC0006619, Sept 1st 2011 to August 31st 2014, $180,000, Co-Investigator (PI: Andrew Robertson, Columbia University).

48. *Copernicus: System for Foresight and Understanding from Scientific Exposition*, IARPA, contract number D11PC20155, September 2011 to August 2016, $1,097,420, Principal Investigator.

47. *Probabilistic Alignment and Distributed Analytics*, IARPA/AFRL FA8650-10-C-7060, Oct 1 2010 to Dec 31 2011, $334,537, Principal Investigator.

46. *Biomedical Informatics Training Program (supplement)*, award number NIH LM07443-10S1, 7/1/10-6/30/11, $153,485, Senior Personnel (PI: Pierre Baldi, UC Irvine).

45. *Automating Behavioral Coding via Text-Mining and Speech Signal Processing*, National Institutes of Health, award number R01AA018673, $3.1 million, (UC Irvine portion is $953,952), Sept 1 2010 to August 31 2015, Co-Investigator (PI: David Atkins, University of Washington).

44. *UC Irvine Clinical Translational Science Center*, National Institutes of Health, award number UL1RR031985, $7,075,320 awarded to date, July 1 2010 to March 31st 2015, Senior Personnel (PI: Dan Cooper, UC Irvine).

43. *Scaling Statistical Topic Modeling Algorithms to Massive Data Sets*, Yahoo! Faculty Research (FREP) award, $10,000 gift, May 2010, Principal Investigator.

42. *Scalable Methods for the Analysis of Network-based Data*, Office of Naval Research: Multidisciplinary University Research Initiative (MURI) Award), award number N00014-08-1-1015, $5,381,300, May 1 2008 to April 30 2013, Principal Investigator.

41. *Scaling Statistical Topic Modeling Algorithms to Massive Data Sets*, Google Research Award, $60,000, April 2008, Principal Investigator.

40. *Research in Cyber-Fraud Detection and Prevention*, gift from Experian, Inc., $200,000, February 2008, Co-Principal Investigator with Michael Goodrich.

39. *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, $360,000, Oct 1 2007 to Sept 30 2010, Principal Investigator.

38. *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-Principal Investigator with Andrew McCallum (University of Massachusetts).

37. *Functional Biomedical Informatics Research Network (FBIRN)*, National Institutes of Health, U24RR021992, $23,992,092, from February 8th 2006 to November 30th 2010, Senior Personnel (PI: Steven Potkin, UC Irvine).

36. *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 Magnusdottir, UC Irvine).

35. *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.

34. *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: Sharon Block, UC Irvine).

33. *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, UC Irvine).

32. *National Alliance for Medical Image Computing (NAMIC), National Institutes of Health*, award number NIH U54 EB005149, total UCI award is $609,253 from 9/17/04 to 8/31/06, Co-Investigator (PI: Ron Kikinis, Brigham and Women's Hospital).

31. *Morphometry Biomedical Informatics Research Network (MBIRN)*, National Institutes of Health, U24-RR021382, total UCI award is $579,880 from 9/30/04 to 5/31/06, Co-Investigator (PI: Bruce Rosen, Massachusetts General Hospital).

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, Principal Investigator.

29. *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.

28. *NSF-ITR: Responding to the Unexpected*, Information Technology Research (ITR) program, National Science Foundation, $9,480,928, award number NSF-ITR-0331707, October 1st 2003 to September 30th 2008, Co-Investgator (PI: Sharad Mehrotra, UC Irvine).

27. *NSF-ITR: The OptIPuter*, Information Technology Research (ITR) program, National Science Foundation, award number , $13,500,000, October 1st 2002 to September 30th 2007, Co-Investigator (PI: Larry Smarr, UCSD).

26. *Biomedical Informatics Training Program*, National Institutes of Health and National Library of Medicine, award number T15-LM-07443, $8,840,297, July 1st 2002 to June 30th 2012, Senior Personnel (PI: Pierre Baldi, UC Irvine).

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, UCLA).

24. *Intelligent Time-Series Pattern Matching*, Jet Propulsion Laboratory, June 15th to September 30th 2002, $80,920, Principal Investigator.

23. *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 (PI: Rod Shankle, UC Irvine).

22. *Spatial Data Mining for Massive Scientific Data Sets*, Lawrence Livermore National Laboratory, May 1st 2001 to August 31st 2002, $100,000, Principal Investigator.

21. *IBM Faculty Partnership Award*, gift from IBM Watson Research Center, May 18th 2001, $40,000, Principal Investigator.

20. *Data Mining of Digital Behavior*, NSF-IIS-0083489, Principal Investigator:
    - Original award: September 15th 2001 to August 30th 2004, $425,000.
    - Supplemental award: September 1st 2003 to December 31st 2010, $1,816,750.

19. *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 (PI: Christine McLaren, UC Irvine).

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.

17. *Sequential Data Analysis for Biomedical Applications*, UCI CORCLR Program, July 1 2000 to June 30th 2001, $12,000, Co-Investigator (PI: Christine McLaren, UC Irvine).

16. *Spatio-Temporal Data Mining of Scientific Trajectory Data*, Lawrence Livermore National Laboratory, March 1st to September 30th 2000, $42,937, Principal Investigator.

15. *Research in Data Mining*, gift from Microsoft Research, October 1999, $60,000, Principal Investigator.

14. *Data Mining of Multivariate Time-Series Sensor Data for Semiconductor Manufacturing*, NIST/National Semiconductor corporation, April 1 1999 through Dec 31 2001, $162,000, Principal Investigator.

13. *Clustering of Sequences and Time Series*, HNC Software, Inc, $40,913, January 1 1999 through Dec 31 1999, Principal Investigator.

12. *SGER: An Online Repository of Large Data Sets for Data Mining Research and Experimentation*, National Science Foundation, NSF IIS-9813584, Aug 15, 1998 to January 31, 2000, $99,737, Principal Investigator.

11. *Data Mining of High-Dimensional Structure-Activity Data Sets*, from SmithKline Beecham Research, September 1st 1998 to April 1st 1999, $22,730, Principal Investigator.

10. *Graduate Fellowships in Biomedical Computing*, US Department of Education, $750,000. Sept 1, 1997 to August 31, 2001, Co-Investigator (PI: Lubomir Bic, UC Irvine).

9. *A Distributed Biomedical Computing Laboratory*, National Science Foundation (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. Co-Investigator.

8. *Turbo-Decoding of High Performance Error-Correcting Codes via Belief Propagation*, AFOSR, grant F49620-97-1-0313, May 1 1997 to December 31 1998, $300,000. Co-Investigator (PI: Robert McEliece, Caltech).

7. *Automated Cloud Screening for Remote Exploration and Experimentation (REE) Applications to the Earth Orbiting-1 (EO-1) Satellite and Similar Platforms*, the Jet Propulsion Laboratory, June 16th 1997 to November 15th 1997, $34,601, Principal Investigator.

6. *Exploring QSAR Data using Probabilistic Data Mining*, SmithKline Beecham Research, July 1st to December 31st 1997, $35,048, Principal Investigator.

5. *Probabilistic Knowledge Discovery and Data Mining: An Integrated Approach at the Interface of Computer Science and Statistics*, National Science Foundation (CAREER award), NSF-9703120, September 1st 1997 to August 31st 2001, $304,379, Principal Investigator.

4. *Clustering and Mode Classification of Engineering Time Series Data*, Jet Propulsion Laboratory, June 15th 1996 to October 17th 1996, $34,401, Principal Investigator.

3. *Automated Detection of Natural Features in SAR Images*, Jet Propulsion Laboratory Director's Discretionary Fund, January 1st 1994 to December 31st 1994, $140,000, Co-Investigator with Usama Fayyad (JPL) and Pietro Perona (Caltech).

2. *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.

1. *An Information-Theoretic Approach to Distributed Inference and Learning*, AFOSR, and ONR. Original award AFOSR-90-0199, February 1st 1990 to May 30th 1992, $338,161. Continuation award NOOO14-92-J-1860: July 1st 1992 to March 30th 1995, $394,118. Co-Investigator (PI: Rodney Goodman, Caltech).

Publications List

Books and Conference Proceedings

B5 A. Nicholson and P. Smyth (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 29th Conference*, ISBN 978-0-9749039-9-6, AUAI Press, Corvallis, OR, 2013.

B4 C. Apte, J. Ghosh, P. Smyth (eds.), *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ISBN 978-1-4503-0813-7, ACM Press, New York, NY, 2011.

B3 *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.

B1 *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

J92 , 'A cell-level discriminative neural network model for diagnosis of blood cancers,' E. E. Robles, Y. Jin, P. Smyth, R. H. Scheuermann, J. D. Bui, H-Y. Wang, J. Oak, Y. Qian, *Bioinformatics*, in press, 2023

J91 'Predicting postfire sediment yields of small steep catchments using airborne lidar differencing,' J. J. Guilinger, E. Foufoula-Georgiou, A. B. Gray, J. Randerson, P. Smyth, N. C. Barth, M. L. Goulden, *Geophysical Research Letters*, in press, 2023.

J90 A. Kumar, P. Smyth, M. Steyvers, ‘Differentiating mental models of self and others: a hierarchical framework for knowledge assessment,’ *Psychological Review*, in press, 2023.

J89 P. Le, J. T. Randerson, R. Willett, S. Wright, P. Smyth, C. Guilloteau, A. Mamalakis, E. Foufoula-Georgiou, ‘Climate-driven changes in the predictability of seasonal precipitation,’ *Nature Communications*, 14(1):3822, 2023.

J88 E. Nalisnick, D. Tran, P. Smyth, ‘A brief tour of deep learning from a statistical perspective,’ *Annual Review of Statistics and its Application*, 10:219–246, https://doi.org/10.1146/annurev-statistics-032921-013738, April 2023.

J87 R. Longjohn, P. Smyth, and H. Stern, ‘Likelihood ratios for categorical count data with applications in digital forensics,’ *Law, Probability, and Risk*, 21(2):91–122, https://doi.org/10.1093/lpr/mgac016, December 2022.

J86 H. Tejeda, A. Kumar, P. Smyth, M. Steyvers, ‘AI-assisted decision-making: a cognitive modeling approach to infer latent reliance strategies,’ *Computational Brain and Behavior*, 5:491-508, https://doi.org/10.1007/s42113-022-00157-y, 2022.

J85 Y. Chen, S. Hantson, N. Andela, S. Coffield, C. Graff, D. Morton, L. Ott, E. Foufoula-Georgiou, P. Smyth, M. Goulden, J. Randerson, ‘California wildfire spread derived using VIIRS satellite observations and an object-based tracking system,’ *Scientific Data*, 9:249, https://doi.org/10.1038/s41597-022-01343-0, 2022.

J84 M. Steyvers, H. Tejeda, G. Kerrigan, P. Smyth, ‘Bayesian modeling of human-AI complementarity,’ *Proceedings of the National Academy of Sciences*, 119(11):1-7, https://doi.org/10.1073/pnas.2111547119, March 2022

J83 H. Do, S. Nandi, P. Putzel, P. Smyth, J. Zhong, ‘A joint fairness model with applications to risk predictions for under-represented populations,’ *Biometrics*, 79(2), 826–840, doi.org/10.1111/biom.13632, published online February 2022 (in print, June 2023).

J82 A. Mamalakis, J. T. Randerson, J-Y Yu, M. Pritchard, G. Magnusdottir, P. Smyth, P. A. Levine, S. Yu, E. Foufoula-Georgiou, ‘Zonally contrasting shifts of the tropical rainbelt in response to climate change,’ *Nature Climate Change*, https://doi.org/10.1038/s41558-020-00963-x, 11: 143151, January 2021.

J81 Park, J., Jindal, A., Kuo, P., Tanana, M., Elston Lafata, J., Tai-Seale, M., Atkins, D. C., Imel, Z. E., Smyth, P, ‘Automated rating of patient and physician emotion in primary care visits,’ *Patient Education and Counseling*, https://doi.org/10.1016/j.pec.2021.01.004, 2021.

J80 A, Stevens, R. Willett, A. Mamalakis, E. Foufoula-Georgiou, A. Tejedor, J. Randerson; P. Smyth, S. Wright., ‘Graph-guided regularized regression of Pacific Ocean climate variables to increase predictive skill of southwestern US winter precipitation,’ *Journal of Climate*, 34(2):737–754, https://doi.org/10.1175/JCLI-D-20-0079.1, 2021.

J79 Y. Chen, J. T. Randerson, S. R. Coffield, E. Foufoula-Georgiou, P. Smyth, C. A. Graff, D. C. Morton, N. Andela, G. R. van der Werf, L. Giglio, L. E. Ott, ‘Forecasting global fire emissions on sub-seasonal to seasonal (S2S) timescales,’ *Journal of Advances in Modeling Earth Systems*, 12(9), e2019MS001955, doi:/10.1029/2019MS001955, 2020.

J78 C. Galbraith, P. Smyth, H. S. Stern, ‘Statistical methods for the forensic analysis of geolocated event data,’ *Forensic Science International*, https://doi.org/10.1016/j.fsidi.2020.301009, 33:1–12, July 2020.

J77 C. Galbraith, P. Smyth, H. Stern, ’Quantifying the association between discrete event time series with applications to digital forensics,’ *Journal of the Royal Statistical Society A*, 183(3):1005–1027, 2020.

J76 C. A. Graff, S. R. Coffield, Y. Chen, E. Foufoula-Georgiou, J. T. Randerson, P. Smyth, ‘Forecasting daily wildfire activity using Poisson regression,’ *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4837–4851, 2020.

J75 R. Baker, D. Xu, J. Park, R. Yu, Q. Li, B. Cung, C. Fischer, F. Rodriguez, M. Warschauer, P. Smyth, 'The benefits and caveats of clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes,' *International Journal of Educational Technology in Higher Education*, 17(13):1–24, 2020.

J74 D. Ji, P. Putzel, Y. Qian, I. Chang, A. Mandava, R. H. Scheuermann, J. D. Bui, H-Y Wang, P. Smyth, 'Machine learning of discriminative gate locations for clinical diagnosis,' *Cytometry A: Special Issue: Machine Learning for Single Cell Data*, 97(3):296–307, 2020.

J73 C. Fischer, Z. Pardos, R. Baker, J. J. Williams, P. Smyth, R. Yu, S. Slater, R. Baker, M. Warschauer, Mining big data in education: Affordances and challenges, *Review of Research in Education*, 44(1):130-160, 2020.

J72 S. Coffield, C. Graff, Y. Chen, P. Smyth, E. Foufoula-Georgiou, J. Randerson, 'Machine learning to predict final fire size at the time of ignition,' *International Journal of Wildland Fire*, 28(11):861–873, 2019.

J71 J. Park, D. Kotzias, P. Kuo, R. L. Logan, K. Merced, S. Singh, M. Tanana, E. Karra-Taniskidou, J. Elston Lafata, D. C. Atkins, M. Tai-Seale, Z. E. Imel, and P. Smyth, 'Detecting conversation topics in primary care office visits from transcripts of patient-provider interactions,' *Journal of the American Medical Informatics Association (JAMIA)*, 26(12):1493–1504, 2019.

J70 D. Kotzias, M. Lichman, and P. Smyth, 'Predicting consumption patterns with repeated and novel events,' *IEEE Transactions on Knowledge and Data Engineering*, 31(2), 371-384, 2018.

J69 J. R. Hipp, C. Bates, M. Lichman, and P. Smyth, 'Using social media to measure temporal ambient population: does it help explain local crime rates?' *Justice Quarterly*, 36(4), 714-748, March 2018.

J68 C. Galbraith and P. Smyth, 'Analyzing user-event data using score-based likelihood ratios with marked point processes,' *Journal of Digital Investigation*, 22, 106-114, 2017.

J67 T. Holsclaw, A. M. Greene, A. W. Robertson, P. Smyth, 'Bayesian non-homogeneous Markov models via Polya-Gamma data augmentation with applications to rainfall modeling', *Annals of Applied Statistics*, 11(1):393–426, 2017.

J66 G. Gaut, M. Steyvers, Z. E. Imel, D. C. Atkins, P. Smyth, 'Content coding of psychotherapy transcripts using labeled topic models,' *IEEE Journal of Biomedical and Health Informatics*, 21(2):476–487, 2017.

J65 C. Haffke, G. Magnusdottir, D. Henke, P. Smyth, Y. Peings, 'Daily states of the March-April east Pacific ITCZ in three decades of high-resolution satellite data,' *Journal of Climate*, doi:10.1175/JCLI-D-15-0224.1, 29(8):2981-2995, 2016.

J64 P. Arnesen, T. Holsclaw, P. Smyth, 'Bayesian detection of changepoints in finite-state Markov chains for multiple sequences,' *Technometrics*, doi:10.1080/00401706.2015.1044118, 58(2), 205-213, 2016.

J63 T. Hoslclaw, A. Greene, A. R. Robertson, P. Smyth, 'A Bayesian hidden Markov model of daily precipitation over South and East Asia,' *Journal of Hydrometeorology*, doi:10.1175/JHM-D-14-0142.1, 17(1):3–25, 2016.

J62 T. Hoslclaw, K. A. Hallgren, M. Steyvers, P. Smyth, D. C. Atkins, 'Measurement error and outcome distributions: Methodological issues in regression analyses of behavioral coding data,' *Psychology of Addictive Behaviors*, doi:10.1037/adb0000091, 29(4):1031-1040, 2015

J61 M. L. Salmans, Z. Yu, K. Watanabe, E. Cam, P. Sun, P. Smyth, X. Dai, B. Andersen, 'The co-factor of LIM domains (CLIM/LDB/NLI) maintains basal mammary epithelial stem cells and promotes breast tumorigenesis,' *PLOS Genetics*, July 2014, doi: 10.1371/journal.pgen.100452.

J60 A. J. Frank, P. Smyth, A. T. Ihler, 'Beyond MAP estimation with the track-oriented multiple hypothesis tracker,' *IEEE Transactions on Signal Processing*, 62(9):2413–2423, 2014.

J59 D. C. Atkins, M. Steyvers, Z. E. Imel, P. Smyth, ‘Scaling up the evaluation of psychotherapy: evaluating motivational interviewing fidelity via statistical text classification,’ *Implementation Science*, 9:49:1–11, 2014.

J58 C. DuBois, C. T. Butts, D. McFarland, P. Smyth, ‘Hierarchical models for relational event sequences,’ *Journal of Mathematical Psychology*, 57(6):297–309, 2013.

J57 N. Navaroli, C. DuBois, P. Smyth, ‘Modeling individual email patterns over time with latent variable models,’ *Machine Learning*, 92(2–3):431-455, May 2013.

J56 M. Geyfman, V. Kumar, Q. Liu, R. Ruiz, W. Gordon, F. Espitia, E. Cam, S. E. Millar, P. Smyth, A. Ihler, J. Takahashi, B. Andersen, ‘Bmal1 controls circadian cell proliferation and susceptibility to UVB-induced DNA damage in the epidermis,’ *Proceedings of the National Academies of Science*, 109(29):11758-63, doi:10.1073/pnas.1209592109, July 2012.

J55 D. Henke, P. Smyth, C. Haffke, G. Magnusdottir, ‘Automated analysis of the temporal behavior of the double Intertropical Convergence Zone over the east Pacific,’ *Remote Sensing of Environment*, 123:418–433, August 2012.

J54 T. Rubin, A. Chambers, P. Smyth, and M. Steyvers, ‘Statistical topic models for multi-label document classification,’ *Machine Learning*, doi: 10.1007/s10994-011-5272-5, 88(1-2):157–208, July 2012.

J53 B. Gretarsson, J. O’ Donovan, S. Bostandjiev, T. Hollerer, A. Asuncion, D. Newman, and P. Smyth, ‘TopicNets: Visual analysis of large text corpora with topic modeling,’ *ACM Transactions on Intelligent Systems and Technology*, 3(2):1–26, February 2012.

J52 M. Steyvers, P. Smyth, and C. Chemudugunta, ‘Combining background knowledge and learned topics,’ *Topics in Cognitive Science*, 3(1):18–47, January 2011.

J51 A. M. Greene, A. W. Robertson, P. Smyth, and S. Triglia, ’Downscaling projections of Indian monsoon rainfall using a nonhomogeneous hidden Markov model,’ *Quarterly Journal of the Royal Meteorological Society*, 137(655):347–359, January 2011.

J50 T. T. Van Leeuwen, A. J. Frank, Y. Jin, P. Smyth, M. L. Goulden, G. R. van der Werf, J. T. Randerson, ’Optimal use of land surface temperature data to detect changes in tropical forest cover,’ *Journal of Geophysical Research—Biogeosciences*, 116, G02002, doi:10.1029/2010JG00148, 2011.

J49 A. Asuncion, P. Smyth, and M. Welling, ’Asynchronous distributed estimation of topic models for document analysis,’ *Statistical Methodology*, 8(1):3–17, January 2011.

J48 C. Bain, G. Magnusdottir, P. Smyth, H. Stern, ‘The diurnal cycle of the intertropical convergence zone in the east Pacific,’ *Journal of Geophysical Research*, 115, D23116, doi:10.1029/2010JD014835, 2010.

J47 C. Bain, J. DePaz, J. Kramer, G. Magnusdottir, P. Smyth, H. Stern, C-C. Wang, ’Detecting the ITCZ in instantaneous satellite data using spatial-temporal statistical modeling: ITCZ climatology in the east Pacific,’ *Journal of Climate*, 138(6):2132-2148, 2010.

J46 S. Kim, P. Smyth, and H. Stern, ’A Bayesian mixture approach to modeling spatial activation patterns in multi-site fMRI data,’ *IEEE Transactions on Medical Imaging*, 29(6):1260–1274, June 2010.

J45 L. Scharenbroich, G. Magnusdottir, P. Smyth, H. Stern and C. Wang, ‘A Bayesian framework for storm tracking using a hidden-state representation,’ *Monthly Weather Review*, 138(6):2132–2148, June 2010.

J44 Q. Liu, K. K. Lin, B. Andersen, P. Smyth, and A. Ihler, ’Estimating replicate time-shifts using Gaussian process regression,’ *Bioinformatics*, 26(6):770–776, 2010.

J43 M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth, and M. Steyvers, ‘Learning author-topic models from text corpora,’ *ACM Transactions on Information Systems*, 28(1):1–38, 2010.

J42 D. Chudova, A. T. Ihler, K. K. Lin, B. Andersen, P. Smyth, 'Bayesian detection of non-sinusoidal periodic patterns in circadian expression data,' *Bioinformatics*, 25(23):3114–3120, 2009.

J41 D. Newman, A. Asuncion, P. Smyth, and M. Welling, 'Distributed algorithms for topic models,' *Journal of Machine Learning Research*, 10:1801–1828, 2009.

J40 K. K. Lin, V. Kumar, M. Geyfman, D. Chudova, A. T. Ihler, P. Smyth, R. Paus, J. S. Takahashi, B. Andersen, 'Circadian clock genes contribute to the regulation of hair follicle cycling,' *PLOS Genetics*, 5(7): e1000573. doi:10.1371/journal.pgen.1000573, 2009.

J39 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):1–23, 2007.

J38 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

J37 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.

J36 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.

J35 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*, 29(8):958–972, 2008.

J34 A. Ihler, S. Kirshner, M. Ghil, A. Robertson, P. Smyth, 'Graphical models for statistical inference and data assimilation,' *Physica D*, 230(1–2):72–87, 2007.

J33 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.

J32 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.

J31 Turner, J.A., Smyth, P., Macciardi, F., Fallon, J.H., Kennedy, J.L., Potkin, S.G., 'Imaging phenotypes and genotypes in schizophrenia,' *Neuroinformatics*, 4(1):21–50, March 2006.

J30 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.

J29 K. K. Lin, D. Chudova, G. W. Hatfield, 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.

J28 D. Pavlov, H. Mannila, and P. Smyth, 'Beyond independence: probabilistic models for query approximation on binary transaction data,' *IEEE Transactions on Knowledge and Data Engineering*, 15(6):1409–1421, September 2003.

J27 I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White, 'Model-based clustering and visualization of navigation patterns on a Web site', *Journal of Data Mining and Knowledge Discovery*, 7(4):399–424, 2003.

J26 D. Chudova and P. Smyth, 'Analysis of pattern discovery in sequences using a Bayes error rate framework,' *Journal of Data Mining and Knowledge Discovery*, 7(3):273–299, 2003.

J25 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.

J24 X. Ge, D. Eppstein, and P. Smyth, 'The distribution of cycle lengths in graphical models for iterative decoding' *IEEE Transactions on Information Theory*, 47(6):2549–2552, September 2001.

J23 P. Smyth, 'Data mining: data analysis on a grand scale?", *Statistical Methods in Medical Research*, 9:309–327, 2000.

J22 P. Smyth 'Model selection for probabilistic clustering using cross-validated likelihood,' *Statistics and Computing*, 9:63–72, 2000.

J21 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.

J20 P. Smyth, K. Ide, and M. Ghil, 'Multiple regimes in Northern hemisphere height fields via mixture model clustering,' *Journal of the Atmospheric Sciences*, 56(21):3704–3723, 1999.

J19 P. Smyth and D. Wolpert, 'Linearly combining density estimators via stacking,' *Machine Learning*, 36(1-2):59–83, July 1999.

J18 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.

J17 P. Smyth, 'Belief networks, hidden Markov models, and Markov random fields: a unifying view,' *Pattern Recognition Letters*, 18:1261–1268, 1997.

J16 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.

J15 C. Brodley and P. Smyth, 'Applying classification algorithms in practice,' *Statistics and Computing*, 7(1):45–56, March 1997.

J14 P. Smyth, D. Heckerman, M. Jordan, 'Probabilistic independence networks for hidden Markov probability models,' *Neural Computation*, 9(2):227–269, 1997.

J13 P. Smyth, 'Bounds on the mean classification error rate of multiple experts,' *Pattern Recognition Letters*, 17:1253–1257, 1996.

J12 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.

J11 P. Smyth, 'Markov monitoring with unknown states,' *IEEE Journal on Selected Areas in Communications*, special issue on Intelligent Signal Processing for Communications, 12(9):1600–1612, December 1994.

J10 A. Y. Lee and P. Smyth, 'Synthesis of minumum-time nonlinear feedback laws for dynamic systems using neural networks,' *Journal of Guidance and Control*, 17(4):868–870, 1994.

J9 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*, 5(2):320–330, March 1994.

J8 P. Smyth, 'Hidden Markov models for fault detection in dynamic systems,' *Pattern Recognition*, 27(1):149–164, 1994.

J7 Z. Zheng, R. Goodman, and P. Smyth, 'Learning finite-state machines with self-clustering recurrent networks,' *Neural Computation*, 5(6):976–990, November 1993.

J6 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*, 39(4):1404–1408, July 1993.

J5 P. Smyth, ‘Admissible stochastic complexity models for classification problems,’ *Statistics and Computing*, 2:97–104, 1992.

J4 R. M. Goodman, P. Smyth, C. Higgins and J. Miller, ‘Rule-based networks for classification and probability estimation,’ *Neural Computation*, 4:781-804, 1992.

J3 P. Smyth and R. M. Goodman, ‘An information theoretic approach to rule induction from databases,’ *IEEE Transactions on Knowledge and Data Engineering*, 4(4):301–316, August 1992.

J2 R. M. Goodman and P. Smyth, ‘Decision tree design using information theory,’ *Knowledge Acquisition*, 4(1):1–26, 1990.

J1 R. M. Goodman and P. Smyth, ‘Decision tree design from a communication theory standpoint,’ *IEEE Transactions on Information Theory*,34(5):979-994, September 1988.

Conference Papers

C131 A. Li, C. Qiu, M. Kloft, P. Smyth, M. Rudoplh, S. Mandt, ‘Batch normalization enables zero-shot anomaly detection,’ *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, Dec 2023, to appear.

C130 H. Do, Y. Chang, Y. S. Cho, P. Smyth, J. Zhong, ’Fair survival time prediction via mutual information minimization,’ *Proceedings of the Machine Learning for Healthcare Conference (MLHC 2023)*, August 2023.

C129 A. Boyd, Y. Chang, S. Mandt, P. Smyth, ‘Inference for mark-censored temporal point processes,’ *Proceedings of the 39th Conference on Uncertainty in AI*, PMLR 216:226236, August 2023.

C128 A. Li, C. Qiu, M. Kloft, P. Smyth, S. Mandt, M. Rudolph, ‘Deep anomaly detection under labeling budget constraints,’ *Proceedings of the 40th International Conference on Machine Learning (ICML 2023)*, PMLR 202:19882-19910, July 2023.

C127 M. Kelly, P. Smyth, M. Steyvers, ‘Capturing humans mental models of AI: an item response theory approach,’ *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*, ACM Press, pp. 1723-1734, https://doi.org/10.1145/3593013.3594111, June 2023.

C126 G. Kerrigan, J. Ley, P. Smyth, ‘Diffusion generative models in infinite dimensions,’ *Proceedings of the 26th International Conference on AI and Statistics*, Proceedings of Machine Learning Research, PMLR 206:9538–9563, April 2023.

C125 A. Boyd, Y. Chang, S. Mandt, P. Smyth, ‘Probabilistic querying of continuous-time event sequences,’ *Proceedings of the 26th International Conference on AI and Statistics*, Proceedings of Machine Learning Research, PMLR 206:10235–10251, April 2023.

C124 M. Kelly, P. Smyth, ‘Variable-based calibration for machine learning classifiers,’ *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, AAAI Press, pp.8211-8217, 2023.

C123 A. Boyd, S. Showalter, S. Mandt, P. Smyth, Predictive querying for autoregressive neural sequence models, *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, 35:23751–23764, 2022.

C122 H. Do, P. Putzel, A. Martin, P. Smyth, J. Zhong, ‘Fair generalized linear models with a convex penalty,’ *Proceedings of the 39th International Conference on Machine Learning (ICML)*, PMLR 162:5286–5308, July 2022.

C121 G. Kerrigan, P. Smyth, and M. Steyvers ‘Combining human predictions with model probabilities via confusion matrices and calibration,’ *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, pp.4421–4434, 2021.

C120 A. Li, A. Boyd, P. Smyth, S. Mandt, 'Detecting and adapting to irregular distribution shifts in Bayesian online learning,' *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, pp.6816–6828, 2021.

C119 P. Putzel, H. Do, P. Smyth, J. Zhong, 'Dynamic survival analysis for EHR data with personalized parametric distributions,' *Proceedings of the 2021 Machine Learning for Healthcare Conference (MLHC)*, PMLR 149:648–673, August 2021.

C118 D. Ji, R. Logan, P. Smyth, and M. Steyvers, 'Active Bayesian assessment for black-box classifiers,' *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI 2021)*, 35(9): 7935–7944, AAAI Press, 2021.

C117 D. Ji, P. Smyth, and M. Steyvers 'Can I trust my fairness metric? Assessing fairness with unlabeled data and Bayesian inference,' *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, pp. 18600–18612, 2020.

C116 A. Boyd, R. Bamler, S. Mandt, and P. Smyth, 'User-dependent neural sequence models for continuous-time event data,' *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, pp. 21488–21499, 2020.

C115 E. Nalisnick, J. M. Hernandez-Lobato, P. Smyth, 'Dropout as a structured shrinkage prior,' *Proceedings of the 36th International Conference on Machine Learning (ICML)*, PMLR 97:4712-4722, 2019.

C114 D. Ji, E. Nalisnick, Y. Qian, R. Scheuermann, P. Smyth, 'Bayesian trees for automated cytometry data analysis,' *Machine Learning for Healthcare (MLHC) 2018 Conference: Proceedings of Machine Learning Research*, PMLR 85:1–18, 2018.

C113 J. Park, R. Yu, F. Rodriguez, R. Baker, P. Smyth, M. Warschauer, 'Understanding student procrastination via mixture models,' *Proceedings of the 2018 Educational Data Mining Conference*, Buffalo, NY, ACM Press, pp.187–197, July 2018 (Best Paper Award).

C112 E. Nalisnick and P. Smyth, 'Learning priors for invariance,' *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics: Proceedings of Machine Learning Research (84)*, pp.366–375, April 2018.

C111 M. Lichman and P. Smyth, 'Prediction of sparse user-item consumption rates with zero-inflated Poisson regression,' *Proceedings of the WWW 2018 Conference*, pp.719–728, ACM Press, April 2018

C110 E. Nalisnick and P. Smyth, 'Learning approximately objective priors,' *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI 2017)*, Association for Uncertainty in Artificial Intelligence (AUAI), August 2017.

C109 E. Nalisnick and P. Smyth, 'Stick-breaking variational autoencoders,' *Proceedings of the International Conference on Learning Representations (ICLR 2017)*, April 2017.

C108 J. Park, K. Denaro, F. Rodriguez, P. Smyth, M. Warschauer, 'Detecting changes in student behavior from clickstream data,' *Proceedings of the Learning Analytics and Knowledge (LAK) Conference*, ACM Press, pp. 21–30, March 2017 (Honorable Mention for Best Paper).

C107 M. Lichman, D. Kotzias, and P. Smyth, 'Personalized location models with adaptive mixtures,' *Proceedings of the ACM SIGSPATIAL Conference*, New York: ACM Press, October 2016.

C106 D. Kotzias, M. Denil, N. De Freitas, P. Smyth, 'From group to individual labels using deep features,' *Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp. 597–606, August 2015.

C105 M. Tanana, K. Hallgren, Z. Imel, D. Atkins, P. Smyth, V. Srikumar, 'Recursive neural networks for coding therapist and patient behavior in motivational interviewing,' in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Association of Computational Linguistics, pp 71–79, 2015.

C104 N. Navaroli and P. Smyth, ‘Modeling response time in digital human communication,’ *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM-2015)*, AAAI Press, pp.278–287, May 2015.

C103 K. Bache, P. Smyth, and D. DeCoste, ‘Hot swapping for online adaptation of optimization hyperparameters,’ *International Conference on Learning Representations (ICLR-2015)*, May 2015.

C102 M. Lichman and P. Smyth, 'Modeling human location data with mixtures of kernel densities,’ *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp. 35-44, August 2014.

C101 J. Foulds and P. Smyth, ‘Annealing paths for the evaluation of topic models,’ *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, AUAI Press: Corvallis, Oregon, pp.220–229, 2014.

C100 C. DuBois, A. Korattika, M. Welling, and P. Smyth, ‘Approximate slice sampling for Bayesian posterior inference,’ in *Proceedings of the 17th International Conference on AI and Statistics*, JMLR Workshop and Conference Proceedings, 33:185–193, 2014.

C99 J. Foulds and P. Smyth, ‘Modeling scientific impact with topical influence regression,’ *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*, Association for Computational Linguistics, pp.113-123, October 2013.

C98 R. Krestel and P. Smyth, ‘Recommending patents based on latent topics,’ *Proceedings of the 7th ACM Recommender Systems Conference (ACM RecSys)*, New York: ACM Press, pp. 395–398, October 2013.

C97 J. Foulds, L. Boyles, C. DuBois, P. Smyth, M. Welling, ‘Stochastic collapsed variational Bayesian inference for latent Dirichlet allocation,’ *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp.446–454, August 2013.

C96 K. Bache, D. Newman, P. Smyth, ‘Text-based measures of topic diversity,’ *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp.23–31, August 2013.

C95 C. DuBois, C. T. Butts, P. Smyth, ‘Stochastic blockmodeling of relational event dynamics,’ in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, Carvalho, Carlos M. and Ravikumar, Pradeep (eds.), JMLR Workshop and Conference Proceedings, 31:238-246, May 2013.

C94 M. J. Bannister, C. DuBois, D. Eppstein, P. Smyth, ‘Windows into relational events: data structures for contiguous subsequences of edges,’ *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, SIAM, pp.856–864, January 2013.

C93 N. Navaroli, C. DuBois, P. Smyth, ‘Statistical models for exploring individual email communication behavior,’ *Proceedings of the 4th Asian Conference on Machine Learning (ACML 2012), JMLR Workshop and Conference Proceedings*, 25:317-332, 2012.

C92 A. Frank, P. Smyth, and A. T. Ihler, ‘A graphical model representation of the track-oriented multiple hypothesis tracker,’ *Proceedings of the IEEE Statistical Signal Processing (SSP) Workshop*, pp.768–771, 2012.

C91 J. Ion Titapiccolo, M. Ferrario, C. Barbieri, D. Marcelli, F. Mari, E. Gatti, S. Ceruto, P. Smyth, and M. G. Signorini, ‘Predictive modeling of cardiovascular complications in incident hemodialysis patients,’ *Proceedings of the IEEE International Conference on Engineering in Medicine and Biology*, IEEE Press, pp.3943–3946, August 2012.

C90 D. Q. Vu, A. Asuncion, D. R. Hunter, and P. Smyth, ‘Continuous-time regression models for longitudinal networks,’ *Proceedings of the 25th Conference on Neural Information Processing Systems (NIPS 2011)*, pp.2492–2500, Dec 2011.

C89 C. DuBois, J. Foulds, and P. Smyth, ‘Latent set models for two-mode data,’ *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, AAAI Press, pp.137-144, 2011.

C88 D. Q. Vu, A. Asuncion, D. R. Hunter, and P. Smyth, ‘Dynamic egocentric models for citation networks,’ *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, Omnipress Conference Publishers, pp.857-864, June 2011.

C87 J. Foulds ,C. DuBois, A. Asuncion, C. T. Butts, and P. Smyth, ‘A dynamic relational infinite feature model for longitudinal social networks,’ *Proceedings of the 14th International Conference on AI and Statistics*, in volume 15 of the *Journal of Machine Learning Research*, 15:287-295, April 2011.

C86 J. Foulds, N. Navaroli, P. Smyth, and A. T. Ihler, ’Revisiting MAP estimation, message passing, and perfect graphs,’ *Proceedings of the 14th International Conference on AI and Statistics*, in volume 15 of the *Journal of Machine Learning Research*, 15:278-286, April 2011.

C85 J. Foulds and P. Smyth, ‘Multi-instance mixture models and semi-supervised learning,’ *Proceedings of the 11th SIAM International Conference on Data Mining*, pp.606–617, April 2011.

C84 A. Chambers, P. Smyth, and M. Steyvers, ‘Learning concept graphs with stick-breaking priors,’ *Neural Information Processing Systems (NIPS) 23*, Cambridge, MA: MIT Press, pp.334–342, December 2010.

C83 C. DuBois and P. Smyth, ’Modeling relational events via latent classes,’ *Proceedings of the Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp.803–812, July 2010.

C82 A. Asuncion, Q. Liu, A. Ihler, and P. Smyth, ‘Particle filtered MCMC-MLE with connections to contrastive divergence,’ *27th International Conference on Machine Learning (ICML 2010)*, Omnipress Conference Publishers, pp.47–54, July 2010.

C81 A. Asuncion, Q. Liu, A. Ihler, and P. Smyth, ‘Learning with blocks: composite likelihood and contrastive divergence,’ *13th International Conference on AI and Statistics*, *JMLR Conference Proceedings*, 9:33-40, May 2010.

C80 A. Ihler, A. J. Frank, P. Smyth, ‘Particle-based variational inference for continuous systems,’ *Neural Information Processing Systems (NIPS) 22*, Cambridge, MA: MIT Press, pp.826–834, December 2009.

C79 A. Asuncion, M. Welling, P. Smyth, and Y. Teh ‘On smoothing and inference for topic models,’ *Proceedings of the 25th Conference on Uncertainty in AI*, AUAI Press, pp.27–34, June 2009.

C78 A. Asuncion, P. Smyth, and M. Welling, ‘Asynchronous distributed learning of topic models,’ *Neural Information Processing Systems (NIPS) 21*, Cambridge, MA: MIT Press, pp.81–88, December 2008.

C77 C. Chemudugunta, P. Smyth, and M. Steyvers, ‘Combining concept hierarchies and statistical topic models,’ *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM-08)*, New York: ACM Press, pp.1469-1470, October 2008.

C76 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.

C75 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.

C74 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.

C73 D. Newman, A. Asuncion, P. Smyth, M. Welling, 'Distributed inference for latent Dirichlet allocation', *Advances in Neural Information Processing Systems 20*, Cambridge MA: MIT Press, pp.1081–1088, December 2007.

C72 S. Kirshner and P. Smyth, 'Infinite mixtures of trees,' in *Proceedings of the 24th International Conference on Machine Learning*, New York: ACM International Conference Proceeding Series, pp.417–723, June 2007.

C71 D. Newman, K. Hagedorn, C. Chemudugunta, and P. Smyth, 'Subject metadata enrichment using statistical topic models,' in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, New York: ACM Press, pp.366–375, June 2007.

C70 A. Ihler and P. Smyth, Learning time-intensity profiles of human activity using non-parametric Bayesian models, *Advances in Neural Information Processing Systems 19*, Cambridge MA: MIT Press, pp.625–632, December 2006.

C69 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*, Cambridge MA: MIT Press, pp. 241-248, 2006.

C68 S. Kim and P. Smyth, Hierarchical Dirichlet processes with random effects, *Advances in Neural Information Processing Systems 19*, Cambridge MA: MIT Press, pp.697–704. December 2006.

C67 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.

C66 D. Newman, C. Chemudugunta, P. Smyth, and M. Steyvers, Statistical entity-topic models, *Proceedings of the Twelvth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp.680–686, August 2006.

C65 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, pp.207–216, August 2006.

C64 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*, 385–392, July 2006.

C63 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, pp.93–104, May 2006.

C62 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.

C61 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, 76–84, April 2005.

C60 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.

C59 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.

C58 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.

C57 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.

C56 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.

C55 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.

C54 D. Chudova, C. Hart, E. Mjolsness and P. Smyth, 'Gene expression clustering with functional mixture models,' in *Advances in Neural Information Processing 16*, MIT Press, 2004.

C53 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. 274–285. August 2003.

C52 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.

C51 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.

C50 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.

C49 D. Pavlov and P. Smyth, 'Approximate query answering by model averaging,' *Proceedings of the SIAM International Conference on Data Mining*, pp.142–153, April 2003.

C48 S. Gaffney and P. Smyth, 'Curve clustering with random effects regression 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.

C47 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.

C46 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.

C45 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.

C44 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*, 153–162, July 2002 (winner, best research paper award).

C43 S. Scott and P. Smyth, 'The Markov modulated Poisson process and Markov 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, 671–680, 2003.

C42 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.

C41 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.

C40 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.

C39 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.

C38 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).

C37 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.

C36 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.

C35 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.

C34 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.

C33 H. Mannila and P. Smyth, 'Approximate query answering with frequent sets and maximum entropy,' *Proceedings of ICDE 2000*, IEEE Press, 309, February 2000.

C32 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.

C31 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.

C30 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.

C29 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.

C28 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.

C27 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).

C26 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.

C25 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.*

C24 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.

C23 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.

C22 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.

C21 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.

C20 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.

C19 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.

C18 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.

C17 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.

C16 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.

C15 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.

C14 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.

C13 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.

C12 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.

C11 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.

C10 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.

C9 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.

C8 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.

C7 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.

C6 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.

C5 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.

C4 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.

C3 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.

C2 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.

C1 R. M. Goodman and P. Smyth, 'Information-theoretic rule induction,' *Proceedings of the 1988 European Conference on Artificial Intelligence*, Pitman: London, 1988.

C0 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.

Commentaries and Journal Editorials

P. Smyth and C. Elkan, Technical perspective: creativity helps influence prediction precision, *Communications of the ACM*, 53(4):88, 2010.

P. Smyth and S. Kirshner, Commentary on article by T. Ryden, *Bayesian Analysis*, (3)4: 699–706, 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 and Opinion Pieces

M13 T. De Bie, L. De Raedt, J. Hernandez-Orallo, H. H. Hoos, P. Smyth, C. K. I. Williams, 'Automating data science: prospects and challenges,' *Communications of the ACM*, 65(3):76–87, March 2022.

M12 Y. Gil et al., 'Intelligent systems for geosciences: an essential research agenda,' *Communications of the ACM*, 62(1):76–84, 2018.

M11 D. Blei and P. Smyth, 'Science and data science,' *Proceedings of the National Academy of Sciences*, 114 (33), 8689–8692, 2017.

M10 P. Smyth and C. Elkan, Technical perspective: Creativity helps influence prediction precision *Communications of the ACM*, 53(4):88, 2010.

M9 J. Bennett, C. Elkan, B. Liu, P. Smyth, D. Tikk, KDD Cup and Workshop 2007, *SIGKDD Explorations*, 9(2), pp.51–52, 2007.

M8 J. O' Madadhain, J. Hutchins, P. Smyth, Prediction and ranking algorithms for event-based network data, *SIGKDD Explorations*, 7(2): 23–30, 2005.

M7 P. Smyth, D. Pregibon, C. Faloutsos, 'Data-driven evolution of data mining algorithms,' *Communications of the ACM*, 45(8), 33–37, August 2002.

M6 C. Apte, B. Liu, E. Pednault, P. Smyth, 'Business applications of data mining,' *Communications of the ACM*, 45(8), 49–53, August 2002.

M5 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.

M4 C. Glymour, D. Madigan, D. Pregibon, and P. Smyth, 'Statistical inference and data mining,' *Communications of the ACM*, 39(11), 35–41,November 1996.

M3 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.

M2 U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, 'From data mining to knowledge discovery,' *AI Magazine*, 37–54, Fall 1996.

M1 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

BC18 A. Greene, T. Holsclaw, A. Robertson, P. Smyth, 'A Bayesian multivariate nonhomogeneous Markov model,' in *Machine Learning and Data Mining Approaches to Climate Science:*, Springer, pp.61–69, 2015.

BC17 A. Asuncion, P. Smyth, M. Welling, D. Newman, I. Porteous, S. Triglia, 'Distributed Gibbs sampling for latent variable models,' in *Scaling Up Machine Learning: Parallel and Distributed Approaches*, R. Bekkerman, M. Bilenko, and J. Langford (eds.), Cambridge University Press, pp. 217–239, 2011.

BC16 J. Hutchins, A. Ihler, and P. Smyth, 'Probabilistic analysis of a large-scale urban traffic data set,' in *Knowledge Discovery from Sensor Data*, N. V. Chawla and A. R. Ganguly (eds.), LNCS 5840, Berlin: Springer Verlag, pp. 94–114, 2010.

BC15 N. Ashish, R. Eguchi, R. Hegde, C. Huyck, D. Kalashnikov, S. Mehrotra, P. Smyth, and N. Venkatasubramanian, Situational awareness technologies for disaster response, in *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*, Chen et al (eds), Springer, pp. 517–544, 2008.

BC14 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.

BC13 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.

BC12 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).

BC11 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.

BC10 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.

BC9 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.

BC7 P. Smyth, 'Learning with probabilistic supervision,' in *Computational Learning Theory and Natural Learning Systems 3*, T. Petcshe, S. Hanson, and J. Shavlik (eds), Cambridge, MA: MIT Press, pp.163–182, 1995.

BC6 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.

BC5 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.

BC4 P. Smyth, ‘Probability density estimation and local basis function neural networks,’ in *Computational Learning Theory and Natural Learning Systems 2*, S. Hanson, T. Petcshe, M. Kearns, R. Rivest (eds), Cambridge, MA: MIT Press, pp.233–248, 1994.

BC3 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).

BC2 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.

BC1 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.

Patents

U. S. Patent no. 4807280, *Cross-Connect Switch*, assigned to Pacific Bell, inventors are E. C. Posner and P. Smyth, issued February 21 1989.

U. S. Patent no. 4845736, *Cross-Connect Switch and Method for Providing Test Access Thereto*, assigned to Pacific Bell, inventors are E. C. Posner and P. Smyth, issued July 4 1989.

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.