Padhraic Smyth

List of Publications by Citations

145 papers	**7,555** citations	**46** h-index	**85** g-index
157 ext. papers	**8,868** ext. citations	**4.1** avg, IF	**6.02** L-index

#	Paper	IF	Citations
145	The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, **1996**, 39, 27-34	2.5	821
144	Probabilistic author-topic models for information discovery **2004**,		273
143	A Spectral Clustering Approach To Finding Communities in Graphs **2005**,		244
142	Fast collapsed gibbs sampling for latent dirichlet allocation **2008**,		239
141	Cluster Analysis of Typhoon Tracks. Part II: Large-Scale Circulation and ENSO. *Journal of Climate*, **2007**, 20, 3654-3676	4.4	211
140	Trajectory clustering with mixtures of regression models **1999**,		210
139	Cluster Analysis of Typhoon Tracks. Part I: General Properties. *Journal of Climate*, **2007**, 20, 3635-3653	4.4	205
138	Test-retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, **2008**, 29, 958-72	5.9	188
137	. *IEEE Transactions on Knowledge and Data Engineering*, **1992**, 4, 301-316	4.2	185
136	Learning author-topic models from text corpora. *ACM Transactions on Information Systems*, **2010**, 28, 1-38	4.8	175
135	Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing*, **2000**, 10, 63-72	1.8	174
134	Brain and muscle Arnt-like protein-1 (BMAL1) controls circadian cell proliferation and susceptibility to UVB-induced DNA damage in the epidermis. *Proceedings of the National Academy of Sciences of the United States of America*, **2012**, 109, 11758-63	11.5	165
133	Statistical topic models for multi-label document classification. *Machine Learning*, **2012**, 88, 157-208	4	162
132	Algorithms for estimating relative importance in networks **2003**,		147
131	Downscaling of Daily Rainfall Occurrence over Northeast Brazil Using a Hidden Markov Model. *Journal of Climate*, **2004**, 17, 4407-4424	4.4	139
130	Multiple Regimes in Northern Hemisphere Height Fields via MixtureModel Clustering*. *Journals of the Atmospheric Sciences*, **1999**, 56, 3704-3723	2.1	136
129	KDD Cup and workshop 2007. *SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining*, **2007**, 9, 51-52	4.6	134

128	Prediction and ranking algorithms for event-based network data. *SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining*, **2005**, 7, 23-30	4.6	131
127	Model-Based Clustering and Visualization of Navigation Patterns on a Web Site. *Data Mining and Knowledge Discovery*, **2003**, 7, 399-424	5.6	127
126	Probabilistic independence networks for hidden Markov probability models. *Neural Computation*, **1997**, 9, 227-69	2.9	122
125	Adaptive event detection with time-varying poisson processes **2006**,		120
124	Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dynamics*, **2007**, 29, 423-440	4.2	119
123	Circadian clock genes contribute to the regulation of hair follicle cycling. *PLoS Genetics*, **2009**, 5, e1000573		117
122	Visualization of navigation patterns on a Web site using model-based clustering **2000**,		104
121	Hidden Markov models for fault detection in dynamic systems. *Pattern Recognition*, **1994**, 27, 149-164	7.7	103
120	Statistical Themes and Lessons for Data Mining. *Data Mining and Knowledge Discovery*, **1997**, 1, 11-28	5.6	97
119	From Group to Individual Labels Using Deep Features **2015**,		96
118	Business applications of data mining. *Communications of the ACM*, **2002**, 45, 49-53	2.5	95
117	The UCI KDD archive of large data sets for data mining research and experimentation. *SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining*, **2000**, 2, 81-85	4.6	94
116	Learning Finite State Machines With Self-Clustering Recurrent Networks. *Neural Computation*, **1993**, 5, 976-990	2.9	86
115	Deformable Markov model templates for time-series pattern matching **2000**,		82
114	Statistical inference and data mining. *Communications of the ACM*, **1996**, 39, 35-41	2.5	79
113	Modeling human location data with mixtures of kernel densities **2014**,		77
112	A general probabilistic framework for clustering individuals and objects **2000**,		77
111	Science and data science. *Proceedings of the National Academy of Sciences of the United States of America*, **2017**, 114, 8689-8692	11.5	76

110	Linearly Combining Density Estimators via Stacking. *Machine Learning*, **1999**, 36, 59-83	4	71
109	Statistical entity-topic models **2006**,		70
108	Identification of hair cycle-associated genes from time-course gene expression profile data by using replicate variance. *Proceedings of the National Academy of Sciences of the United States of America*, **2004**, 101, 15955-60	11.5	69
107	Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, **2020**, 44, 130-160	4.9	64
106	Scaling up the evaluation of psychotherapy: evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science*, **2014**, 9, 49	8.4	62
105	TopicNets. *ACM Transactions on Intelligent Systems and Technology*, **2012**, 3, 1-26	8	62
104	Belief networks, hidden Markov models, and Markov random fields: A unifying view. *Pattern Recognition Letters*, **1997**, 18, 1261-1268	4.7	58
103	Rule-Based Neural Networks for Classification and Probability Estimation. *Neural Computation*, **1992**, 4, 781-804	2.9	50
102	Stochastic collapsed variational Bayesian inference for latent Dirichlet allocation **2013**,		47
101	Learning to Recognize Volcanoes on Venus. *Machine Learning*, **1998**, 30, 165-194	4	46
100	. *IEEE Journal on Selected Areas in Communications*, **1994**, 12, 1600-1612	14.2	46
99	Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland. *Quarterly Journal of the Royal Meteorological Society*, **2006**, 132, 519-542	6.4	45
98	Discrete recurrent neural networks for grammatical inference. *IEEE Transactions on Neural Networks*, **1994**, 5, 320-30		44
97	Using Social Media to Measure Temporal Ambient Population: Does it Help Explain Local Crime Rates?. *Justice Quarterly*, **2019**, 36, 718-748	2.4	44
96	Analyzing Entities and Topics in News Articles Using Statistical Topic Models. *Lecture Notes in Computer Science*, **2006**, 93-104	0.9	40
95	Modeling Documents by Combining Semantic Concepts with Unsupervised Statistical Learning. *Lecture Notes in Computer Science*, **2008**, 229-244	0.9	40
94	. *IEEE Transactions on Information Theory*, **1988**, 34, 979-994	2.8	39
93	Graphical models for statistical inference and data assimilation. *Physica D: Nonlinear Phenomena*, **2007**, 230, 72-87	3.3	38

92	Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. *Quarterly Journal of the Royal Meteorological Society*, **2011**, 137, 347-359	6.4	36
91	Beyond independence: probabilistic models for query approximation on binary transaction data. *IEEE Transactions on Knowledge and Data Engineering*, **2003**, 15, 1409-1421	4.2	36
90	Intelligent systems for geosciences. *Communications of the ACM*, **2018**, 62, 76-84	2.5	34
89	Translation-invariant mixture models for curve clustering **2003**,		31
88	Detecting the ITCZ in Instantaneous Satellite Data using Spatiotemporal Statistical Modeling: ITCZ Climatology in the East Pacific. *Journal of Climate*, **2011**, 24, 216-230	4.4	30
87	Bounds on the mean classification error rate of multiple experts. *Pattern Recognition Letters*, **1996**, 17, 1253-1257	4.7	30
86	Optimal use of land surface temperature data to detect changes in tropical forest cover. *Journal of Geophysical Research*, **2011**, 116,		29
85	Learning to detect events with Markov-modulated poisson processes. *ACM Transactions on Knowledge Discovery From Data*, **2007**, 1, 13	4	29
84	Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction **2001**,		28
83	Hierarchical models for relational event sequences. *Journal of Mathematical Psychology*, **2013**, 57, 297-309	1.2	27
82	Imaging phenotypes and genotypes in schizophrenia. *Neuroinformatics*, **2006**, 4, 21-49	3.2	26
81	Data-driven evolution of data mining algorithms. *Communications of the ACM*, **2002**, 45, 33-37	2.5	25
80	Text-based measures of document diversity **2013**,		24
79	Maximum Likelihood Estimation of Mixture Densities for Binned and Truncated Multivariate Data. *Machine Learning*, **2002**, 47, 7-34	4	24
78	Data mining: data analysis on a grand scale?. *Statistical Methods in Medical Research*, **2000**, 9, 309-27	2.3	24
77	Towards scalable support vector machines using squashing **2000**,		24
76	Zonally contrasting shifts of the tropical rainbelt in response to climate change. *Nature Climate Change*, **2021**, 11, 143-151	21.4	23
75	Applying classification algorithms in practice. *Statistics and Computing*, **1997**, 7, 45-56	1.8	22

74	Detecting changes in student behavior from clickstream data **2017**,		21
73	A Bayesian Hidden Markov Model of Daily Precipitation over South and East Asia. *Journal of Hydrometeorology*, **2016**, 17, 3-25	3.7	20
72	Modeling Count Data from Multiple Sensors: A Building Occupancy Model **2007**,		20
71	Combining concept hierarchies and statistical topic models **2008**,		19
70	Subject metadata enrichment using statistical topic models **2007**,		18
69	The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes. *International Journal of Educational Technology in Higher Education*, **2020**, 17,	6.3	18
68	Content Coding of Psychotherapy Transcripts Using Labeled Topic Models. *IEEE Journal of Biomedical and Health Informatics*, **2017**, 21, 476-487	7.2	17
67	Bayesian nonhomogeneous Markov models via Pólya-Gamma data augmentation with applications to rainfall modeling. *Annals of Applied Statistics*, **2017**, 11,	2.1	17
66	Decision tree design using information theory. *International Journal of Human-Computer Studies*, **1990**, 2, 1-19		17
65	Combining background knowledge and learned topics. *Topics in Cognitive Science*, **2011**, 3, 18-47	2.5	15
64	A Bayesian mixture approach to modeling spatial activation patterns in multisite fMRI data. *IEEE Transactions on Medical Imaging*, **2010**, 29, 1260-74	11.7	15
63	EventRank **2005**,		15
62	Daily States of the March–April East Pacific ITCZ in Three Decades of High-Resolution Satellite Data. *Journal of Climate*, **2016**, 29, 2981-2995	4.4	15
61	Estimating replicate time shifts using Gaussian process regression. *Bioinformatics*, **2010**, 26, 770-6	7.2	14
60	Bayesian detection of non-sinusoidal periodic patterns in circadian expression data. *Bioinformatics*, **2009**, 25, 3114-20	7.2	14
59	Detecting very early stages of dementia from normal aging with Machine Learning methods. *Lecture Notes in Computer Science*, **1997**, 71-85	0.9	14
58	Automated analysis of the temporal behavior of the double Intertropical Convergence Zone over the east Pacific. *Remote Sensing of Environment*, **2012**, 123, 418-433	13.2	12
57	Detecting conversation topics in primary care office visits from transcripts of patient-provider interactions. *Journal of the American Medical Informatics Association: JAMIA*, **2019**, 26, 1493-1504	8.6	11

56	Automated analysis and exploration of image databases: Results, progress, and challenges. *Journal of Intelligent Information Systems*, **1995**, 4, 7-25	2.1	11
55	Measurement error and outcome distributions: Methodological issues in regression analyses of behavioral coding data. *Psychology of Addictive Behaviors*, **2015**, 29, 1031-40	3.4	10
54	The co-factor of LIM domains (CLIM/LDB/NLI) maintains basal mammary epithelial stem cells and promotes breast tumorigenesis. *PLoS Genetics*, **2014**, 10, e1004520	6	10
53	Asynchronous distributed estimation of topic models for document analysis. *Statistical Methodology*, **2011**, 8, 3-17		10
52	A Bayesian Framework for Storm Tracking Using a Hidden-State Representation. *Monthly Weather Review*, **2010**, 138, 2132-2148	2.4	10
51	Predicting Consumption Patterns with Repeated and Novel Events. *IEEE Transactions on Knowledge and Data Engineering*, **2019**, 31, 371-384	4.2	10
50	Machine learning to predict final fire size at the time of ignition. *International Journal of Wildland Fire*, **2019**, 28, 861-873	3.2	9
49	Beyond MAP Estimation With the Track-Oriented Multiple Hypothesis Tracker. *IEEE Transactions on Signal Processing*, **2014**, 62, 2413-2423	4.8	9
48	Recommending patents based on latent topics **2013**,		9
47	A nonparametric bayesian approach to detecting spatial activation patterns in fMRI data. *Lecture Notes in Computer Science*, **2006**, 9, 217-24	0.9	9
46	Modeling relational events via latent classes **2010**,		8
45	Prediction of Sparse User-Item Consumption Rates with Zero-Inflated Poisson Regression **2018**,		8
44	Bayesian Detection of Changepoints in Finite-State Markov Chains for Multiple Sequences. *Technometrics*, **2016**, 58, 205-213	1.4	7
43	Discovering Chinese words from unsegmented text (poster abstract) **1999**,		6
42	Probabilistic Analysis of a Large-Scale Urban Traffic Sensor Data Set. *Lecture Notes in Computer Science*, **2010**, 94-114	0.9	6
41	Forecasting Daily Wildfire Activity Using Poisson Regression. *IEEE Transactions on Geoscience and Remote Sensing*, **2020**, 58, 4837-4851	8.1	5
40	Multi-Instance Mixture Models and Semi-Supervised Learning **2011**,		5
39	Probabilistic query models for transaction data **2001**,		5

38	Graph-Guided Regularized Regression of Pacific Ocean Climate Variables to Increase Predictive Skill of Southwestern U.S. Winter Precipitation. *Journal of Climate*, **2020**, 34, 737-754	4.4	5
37	Technical perspectiveCreativity helps influence prediction precision. *Communications of the ACM*, **2010**, 53, 88-88	2.5	4
36	The distribution of loop lengths in graphical models for turbo decoding. *IEEE Transactions on Information Theory*, **2001**, 47, 2549-2553	2.8	4
35	Cataloging and Mining Massive Datasets for Science Data Analysis. *Journal of Computational and Graphical Statistics*, **1999**, 8, 589-610	1.4	4
34	Cataloging and Mining Massive Datasets for Science Data Analysis. *Journal of Computational and Graphical Statistics*, **1999**, 8, 589	1.4	4
33	Windows into Relational Events: Data Structures for Contiguous Subsequences of Edges **2013**,		4
32	Data Mining at the Interface of Computer Science and Statistics. *Massive Computing*, **2001**, 35-61		4
31	Forecasting Global Fire Emissions on Subseasonal to Seasonal (S2S) Time Scales. *Journal of Advances in Modeling Earth Systems*, **2020**, 12, e2019MS001955	7.1	4
30	Quantifying the association between discrete event time series with applications to digital forensics. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **2020**, 183, 1005-1027	2.1	3
29	A graphical model representation of the track-oriented multiple hypothesis tracker **2012**,		3
28	Synthesis of minimum-time feedback laws for dynamic systems using neural networks. *Journal of Guidance, Control, and Dynamics*, **1994**, 17, 868-870	2.1	3
27	Admissible stochastic complexity models for classification problems. *Statistics and Computing*, **1992**, 2, 97-104	1.8	3
26	Personalized location models with adaptive mixtures **2016**,		3
25	Machine Learning of Discriminative Gate Locations for Clinical Diagnosis. *Cytometry Part A: the Journal of the International Society for Analytical Cytology*, **2020**, 97, 296-307	4.6	3
24	Parametric response surface models for analysis of multi-site fMRI data. *Lecture Notes in Computer Science*, **2005**, 8, 352-9	0.9	3
23	Analyzing user-event data using score-based likelihood ratios with marked point processes. *Digital Investigation*, **2017**, 22, S106-S114	3.3	2
22	Analysis of Pattern Discovery in Sequences Using a Bayes Error Framework. *Data Mining and Knowledge Discovery*, **2003**, 7, 273-299	5.6	2
21	Commerce on the Web: Models and Applications211-234		2

#	Reference		
20	Bayesian modeling of human-AI complementarity.. *Proceedings of the National Academy of Sciences of the United States of America*, **2022**, 119, e2111547119	11.5	2
19	Modeling individual email patterns over time with latent variable models. *Machine Learning*, **2013**, 92, 431-455	4	1
18	Learning with Probabilistic Representations. *Machine Learning*, **1997**, 29, 91-101	4	1
17	Infinite mixtures of trees **2007**,		1
16	Automated Induction of Rule-based Neural Networks from Databases. *Intelligent Systems in Accounting, Finance and Management*, **1993**, 2, 41-54	2.5	1
15	A Bayesian Multivariate Nonhomogeneous Markov Model **2015**, 61-69		1
14	Automated rating of patient and physician emotion in primary care visits. *Patient Education and Counseling*, **2021**, 104, 2098-2105	3.1	1
13	Distributed Gibbs Sampling for Latent Variable Models217-239		
12	Appendix A: Mathematical Complements **2003**, 235-251		
11	Appendix B: List of Main Symbols and Abbreviations **2003**, 253-256		
10	Link Analysis125-147		
9	Advanced Crawling Techniques149-170		
8	Modeling and Understanding Human Behavior on the Web171-209		
7	Text Analysis77-123		
6	Basic WWW Technologies29-50		
5	Web Graphs51-76		
4	Detecting novel fault conditions with hidden Markov models and neural networks. *Machine Intelligence and Pattern Recognition*, **1994**, 16, 525-536		
3	Statistical Methods for the Forensic Analysis of Geolocated Event Data. *Forensic Science International: Digital Investigation*, **2020**, 33, 301009		1

2 Dynamic Survival Analysis with Individualized Truncated Parametric Distributions.. *Proceedings of Machine Learning Research*, **2021**, 146, 159-170   0.4

1 Dynamic Survival Analysis for EHR Data with Personalized Parametric Distributions.. *Proceedings of Machine Learning Research*, **2021**, 149, 648-673   0.4