

REFERENCES

- [1] 2020. Technical Report. <https://www.dropbox.com/sh/ykli4ym6koh8h7v/AADuTh-I9RG15sWuwFfdQyMsa?dl=0>.
- [2] Fatemeh Amiri, Mohammad Mahdi Rezaei Yousefi, Caro Lucas, Azadeh Shakery, and Nasser Yazdani. 2011. Mutual information-based feature selection for intrusion detection systems. *J. Netw. Comput. Appl.* 34, 4 (2011), 1184–1199.
- [3] Bădulescu Laviniu Aurelian. 2018. An information entropy based splitting criterion better for the Data Mining Decision Tree algorithms. In *ICSTCC*. 535–540.
- [4] Daniel Barbará, Yi Li, and Julia Couto. 2002. COOLCAT: an entropy-based algorithm for categorical clustering. In *CIKM*. 582–589.
- [5] Boyan Bonev, Francisco Escolano, and Miguel Cazorla. 2008. Feature selection, mutual information, and the classification of high-dimensional patterns. *Pattern Anal. Appl.* 11, 3-4 (2008), 309–319.
- [6] Wei Cao, Jian Li, Yufei Tao, and Zhize Li. 2015. On Top-k Selection in Multi-Armed Bandits and Hidden Bipartite Graphs. In *NeurIPS*. 1036–1044.
- [7] Graham Cormode. 2013. The continuous distributed monitoring model. *SIGMOD Rec.* 42, 1 (2013), 5–14.
- [8] Graham Cormode and Marios Hadjieleftheriou. 2009. Finding the frequent items in streams of data. *Commun. ACM* 52, 10 (2009), 97–105.
- [9] Graham Cormode and S. Muthukrishnan. 2005. An improved data stream summary: the count-min sketch and its applications. *J. Algorithms* 55, 1 (2005), 58–75.
- [10] Luis M. de Campos. 2006. A Scoring Function for Learning Bayesian Networks based on Mutual Information and Conditional Independence Tests. *J. Mach. Learn. Res.* 7 (2006), 2149–2187.
- [11] Ran El-Yaniv and Dmitry Pechyony. 2009. Transductive Rademacher Complexity and its Applications. *J. Artif. Intell. Res.* 35 (2009), 193–234.
- [12] Pablo A. Estévez, M. Tesmer, Claudio A. Perez, and Jacek M. Zurada. 2009. Normalized Mutual Information Feature Selection. *IEEE Trans. Neural Networks* 20, 2 (2009), 189–201.
- [13] François Fleuret. 2004. Fast Binary Feature Selection with Conditional Mutual Information. *J. Mach. Learn. Res.* 5 (2004), 1531–1555.
- [14] Pawel Foremski, David Plonka, and Arthur W. Berger. 2016. Entropy/IP: Uncovering Structure in IPv6 Addresses. In *IMC*. 167–181.
- [15] Moshe Gabel, Daniel Keren, and Assaf Schuster. 2017. Anarchists, Unite: Practical Entropy Approximation for Distributed Streams. In *SIGKDD*. 837–846.
- [16] Yanjun Han, Jiantao Jiao, and Tsachy Weissman. 2015. Adaptive estimation of Shannon entropy. In *ISIT*. 1372–1376.
- [17] Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. 2015. Minimax Estimation of Functionals of Discrete Distributions. *IEEE Trans. Inf. Theory* 61, 5 (2015), 2835–2885.
- [18] Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. 2017. Maximum Likelihood Estimation of Functionals of Discrete Distributions. *IEEE Trans. Inf. Theory* 63, 10 (2017), 6774–6798.
- [19] Ambika Kaul, Saket Maheshwary, and Vikram Pudi. 2017. AutoLearn - Automated Feature Generation and Selection. In *ICDM*. 217–226.
- [20] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, and Huan Liu. 2018. Feature Selection: A Data Perspective. *ACM Comput. Surv.* 50, 6 (2018), 94:1–94:45.
- [21] Tao Li, Sheng Ma, and Mitsunori Ogihara. 2004. Entropy-based criterion in categorical clustering. In *ICML*, Carla E. Brodley (Ed.), Vol. 69.
- [22] Colin McDiarmid. 1989. On the method of bounded differences. *Surveys in combinatorics* 141, 1 (1989), 148–188.
- [23] Rajeev Motwani and Prabhakar Raghavan. 1995. *Randomized algorithms*. Cambridge university press.
- [24] Jaganathan Palanichamy and Kuppuchamy Ramasamy. 2013. A threshold fuzzy entropy based feature selection for medical database classification. *Comput. Biol. Medicine* 43, 12 (2013), 2222–2229.
- [25] Liam Paninski. 2003. Estimation of Entropy and Mutual Information. *Neural Computation* 15, 6 (2003), 1191–1253.
- [26] Hanchuan Peng, Fuhui Long, and Chris H. Q. Ding. 2005. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 8 (2005), 1226–1238.
- [27] J. Ross Quinlan. 1986. Induction of Decision Trees. *Mach. Learn.* 1, 1 (1986), 81–106.
- [28] Claude E Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.
- [29] Cheng Sheng, Yufei Tao, and Jianzhong Li. 2012. Exact and approximate algorithms for the most connected vertex problem. *ACM Trans. Database Syst.* 37, 2 (2012), 12:1–12:39.
- [30] Paul Valiant and Gregory Valiant. 2013. Estimating the Unseen: Improved Estimators for Entropy and other Properties. In *NeurIPS*. 2157–2165.
- [31] Jorge R. Vergara and Pablo A. Estévez. 2014. A review of feature selection methods based on mutual information. *Neural Computing and Applications* 24, 1 (2014), 175–186.
- [32] Chi Wang and Bailu Ding. 2019. Fast Approximation of Empirical Entropy via Subsampling. In *SIGKDD*. 658–667.
- [33] Qing Ren Wang and Ching Y. Suen. 1984. Analysis and Design of a Decision Tree Based on Entropy Reduction and Its Application to Large Character Set Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 4 (1984), 406–417.
- [34] Sibow Wang, Youze Tang, Xiaokui Xiao, Yin Yang, and Zengxiang Li. 2016. HubPPR: Effective Indexing for Approximate Personalized PageRank. *PVLDB* 10, 3 (2016), 205–216.
- [35] Sibow Wang and Yufei Tao. 2018. Efficient Algorithms for Finding Approximate Heavy Hitters in Personalized PageRanks. In *SIGMOD*. 1113–1127.
- [36] Sibow Wang, Renchi Yang, Runhui Wang, Xiaokui Xiao, Zhewei Wei, Wenqing Lin, Yin Yang, and Nan Tang. 2019. Efficient Algorithms for Approximate Single-Source Personalized PageRank Queries. *ACM Trans. Database Syst.* 44, 4 (2019), 18:1–18:37.
- [37] Sibow Wang, Renchi Yang, Xiaokui Xiao, Zhewei Wei, and Yin Yang. 2017. FORA: Simple and Effective Approximate Single-Source Personalized PageRank. In *SIGKDD*. 505–514.
- [38] Yihong Wu and Pengkun Yang. 2016. Minimax Rates of Entropy Estimation on Large Alphabets via Best Polynomial Approximation. *IEEE Trans. Inf. Theory* 62, 6 (2016), 3702–3720.
- [39] Yiming Yang and Jan O. Pedersen. 1997. A Comparative Study on Feature Selection in Text Categorization. In *ICML*, Douglas H. Fisher (Ed.). Morgan Kaufmann, 412–420.