





















- [17] Hugo Larochelle and Yoshua Bengio. 2008. Classification using discriminative restricted Boltzmann machines. In *ICML (ACM International Conference Proceeding Series)*, Vol. 307. ACM, 536–543.
- [18] Baoli Li and Carl Vogel. 2010. Improving Multiclass Text Classification with Error-Correcting Output Coding and Sub-class Partitions. In *Canadian Conference on AI (Lecture Notes in Computer Science)*, Vol. 6085. Springer, 4–15.
- [19] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning Word Vectors for Sentiment Analysis. In *ACL. The Association for Computer Linguistics*, 142–150.
- [20] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge University Press.
- [21] Hristo S. Paskov, Robert West, John C. Mitchell, and Trevor J. Hastie. 2013. Compressive Feature Learning. In *NIPS*. 2931–2939.
- [22] Jorma Rissanen. 1978. Modeling by shortest data description. *Automatica* 14, 5 (1978), 465–471.
- [23] David Sculley and Carla E. Brodley. 2006. Compression and Machine Learning: A New Perspective on Feature Space Vectors. In *DCC*. IEEE Computer Society, 332–332.
- [24] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural Machine Translation of Rare Words with Subword Units. In *ACL (1)*. The Association for Computer Linguistics.
- [25] Henning Sperr, Jan Niehues, and Alex Waibel. 2013. Letter N-Gram-based Input Encoding for Continuous Space Language Models. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*. Association for Computational Linguistics, 30–39.
- [26] Bo Tang, Steven Kay, and Haibo He. 2016. Toward Optimal Feature Selection in Naive Bayes for Text Categorization. *IEEE Trans. Knowl. Data Eng.* 28, 9 (2016), 2508–2521.
- [27] Alper Kursat Uysal and Serkan Günel. 2012. A novel probabilistic feature selection method for text classification. *Knowl.-Based Syst.* 36 (2012), 226–235.
- [28] 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.
- [29] Xiang Zhang, Junbo Jake Zhao, and Yann LeCun. 2015. Character-level Convolutional Networks for Text Classification. In *NIPS*. 649–657.
- [30] Jacob Ziv and Abraham Lempel. 1977. A universal algorithm for sequential data compression. *IEEE Trans. Information Theory* 23, 3 (1977), 337–343.