

Barbiero, Pietro; Ciravegna, Gabriele; Cirrincione, Giansalvo; Tonda, Alberto; Squillero, Giovanni

Generating neural archetypes to instruct fast and interpretable decisions. (English)

Zbl 1442.62766

Bucciarelli, Edgardo (ed.) et al., Decision economics: complexity of decisions and decisions for complexity. Papers based on the presentations at the international conference on decision economics, DECON 2019, Ávila, Spain, June 26–28, 2019. Cham: Springer. Adv. Intell. Syst. Comput. 1009, 45-52 (2020).

Summary: In the field of artificial intelligence, agents learn how to take decisions by fitting their parameters on a set of samples called training set. Similarly, a core set is a subset of the training samples such that, if an agent exploits this set to fit its parameters instead of the whole training set, then the quality of the inferences does not change significantly. Relaxing the constraint that restricts the search for core sets to the available data, neural networks may be used to generate virtual samples, called archetype set, containing the same kind of information. This work illustrates the features of GH-ARCH, a recently proposed self-organizing hierarchical neural network for archetype discovery. Experiments show how the use of archetypes allows both ML agents to make fast and accurate predictions and human experts to make sense of such decisions by analyzing few important samples.

For the entire collection see [Zbl 1444.91005].

MSC:

- 62R07 Statistical aspects of big data and data science
- 62H30 Classification and discrimination; cluster analysis (statistical aspects)
- 62M45 Neural nets and related approaches to inference from stochastic processes
- 62M20 Inference from stochastic processes and prediction
- 62C25 Compound decision problems in statistical decision theory
- 68T05 Learning and adaptive systems in artificial intelligence

Keywords:

archetypes; big data; classification; coresets; explain AI; GH-ARCH; hierarchical clustering; machine learning; neural networks; self-organization; semi-supervised learning

Software:

Scikit

Full Text: [DOI](#)

References:

- [1] Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608 (2017)
- [2] Mitchell Waldrop, M.: News feature: what are the limits of deep learning? Proc. Natl. Acad. Sci. 116(4), 1074-1077 (2019)
- [3] Bachem, O., Lucic, M., Krause, A.: Practical coresets constructions for machine learning. arXiv preprint arXiv:1703.06476 (2017)
- [4] Ciravegna, G., Barbiero, P., Cirrincione, G., Squillero, G., Tonda, A.: Discovering hierarchical neural archetype sets. In: The International Joint Conference on Neural Networks (IJCNN), July 2019
- [5] Cirrincione, G., Ciravegna, G., Barbiero, P., Randazzo, V., Pasero, E.: The GH-EXIN neural network for hierarchical clustering. Neural Networks (2019, under peer review)
- [6] Haykin, S.S.: Neural Networks and Learning Machines, vol. 3. Pearson, Upper Saddle River (2009)
- [7] Cirrincione, G., Randazzo, V., Pasero, E.: The growing curvilinear component analysis (GCCA) neural network. Neural Netw. 103, 108-117 (2018) · [Zbl 1434.68503](#)
- [8] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825-2830 (2011) · [Zbl 1280.68189](#)

- [9] Breiman, L.: Random forests. *Mach. Learn.* 45(1), 5-32 (2001) · [Zbl 1007.68152](#)
- [10] Tikhonov, A.N.: On the stability of inverse problems. *Dokl. Akad. Nauk SSSR* 39, 195-198 (1943)
- [11] Vanschoren, J., van Rijn, J.N., Bischl, B., Torgo, L.: OpenML: networked science in machine learning. *SIGKDD Explorations* 15(2), 49-60 (2013)
- [12] Breiman, L.: Pasting small votes for classification in large databases and on-line. *Mach. Learn.* 36(1-2), 85-103 (1999)
- [13] Cox, D.R.: The regression analysis of binary sequences. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 20, 215-242 (1958) · [Zbl 0088.35703](#)
- [14] Campbell, T., Broderick, T.: Bayesian coresets construction via greedy iterative geodesic ascent. In: *International Conference on Machine Learning (ICML)* (2018)
- [15] Clarkson, K.L.: Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. *ACM Trans. Algorithms* 6, 63 (2010) · [Zbl 1300.90026](#)
- [16] Pati, Y.C., Rezaifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pp. 40-44 (1993)
- [17] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. *Ann. Stat.* 32(2), 407-451 (2004) · [Zbl 1091.62054](#)
- [18] Boutsidis, C., Drineas, P., Magdon-Ismail, M.: Near-optimal coresets for least-squares regression. *Technical report* (2013) · [Zbl 1364.62170](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.