

Random Forest with Distributed Computing System

ENGR-E599

(Bibliography)

Ronak Shah

ronashah@iu.edu

Shyam Narasimhan

shynaras@iu.edu

- [1] Jeffrey Dean and Sanjay Ghemawat, "Mapreduce: simplified data processing on large clusters," in Proceedings of the 6th conference on Symposium on Operating Systems Design & Implementation -Volume 6, Berkeley, CA, USA, 2004, pp. 10–10, USENIX Association.
- [2] "Apache hadoop," <http://hadoop.apache.org>, Dec 2010.
https://www.researchgate.net/publication/260300430_Parallelizing_Machine_Learning_Algorithms [accessed Dec 09 2018].
- [3] C.T. Chu, S.K. Kim, Y.A. Lin, Y.Y. Yu, G. Bratski, A.Y. Ng, and K. Olukotun, "Map-reduce for machine learning on multi-core," in Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference. The MIT Press, 2007, p. 281.
https://www.researchgate.net/publication/260300430_Parallelizing_Machine_Learning_Algorithms [accessed Dec 09 2018].
- [3] Genuer, R., Poggi, J.-M., Tuleau-Malot, C., Villa-Vialaneix, N., 2015. Random Forests for Big Data. arXiv preprint arXiv:1511.08327.
- [4] Wright, M.N., Ziegler, A., 2015. ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint arXiv:1508.04409.
- [5] Jianguo Chen, Kenli Li, Zhuo Tang : A parallel random forest algorithm for big data in a Spark Cloud Computing Environment
- [6] https://www.researchgate.net/publication/260300430_Parallelizing_Machine_Learning_Algorithms
- [7] <https://www.analyticsvidhya.com/blog/2018/08/dask-big-datasets-machine-learning-python/>
- [8] <https://spark.apache.org/>
- [9] <http://docs.dask.org/en/latest/>
- [10] <http://joblib.readthedocs.io/en/latest/parallel.html>
- [11] <https://github.com/scikit-learn/scikit-learn/blob/master/sklearn/ensemble/forest.py>
- [12] <https://github.com/scikit-learn/scikit-learn/blob/master/sklearn/utils/joblib.py>