DOI: 10.21105/joss.00373

Software

- Review
- Repository
- Archive

array_split: Multi-dimensional array partitioning

Shane J Latham[1]

1 Department of Applied Mathematics, Research School of Physics and Engineering, The Australian National University

Summary

The `array_split` (Latham 2017) Python package extends existing dense array partitioning capabilities found in the `numpy` (Walt, Colbert, and Varoquaux 2011) (`numpy.array_split`) and `skimage` (Van der Walt et al. 2014) (`skimage.util.view_as_blocks`) Python packages. In particular, it provides the means for partitioning based on *array shape* (rather than requiring an actual `numpy.ndarray` object) and can partition into *sub-arrays* based on a variety of criteria including: per-axis number of partitions, total number of sub-arrays (with per-axis number of partition constraints), explicit sub-array shape and constraining a partitioning with an upper bound on the resulting sub-array number of bytes.

Application areas include:

Parallel Processing Data parallelism by partitioning array for multi-process concurrency (e.g. `multiprocessing` ("Multiprocessing – Process-Based Parallelism" 2017) or `mpi4py` (Dalcin et al. 2011)) based on number of cores, or partitioning for accelerator hardware concurrency (e.g. `pyopencl` or `pycuda` [kloeckner_pycuda_2012]) based on hardware memory limits.

File I/O Partitioning large arrays for output to separate files (e.g. as part of a virtual dataset (The HDF Group 1997–1997-NNNN, Collette (2013))) based on maximum file size, or out-of-core partitioning based on in-core memory limits.

References

Collette, Andrew. 2013. *Python and Hdf5*. O'Reilly.

Dalcin, Lisandro D, Rodrigo R Paz, Pablo A Kler, and Alejandro Cosimo. 2011. "Parallel Distributed Computing Using Python." *Advances in Water Resources* 34 (9). Elsevier: 1124–39. doi:10.1016/j.advwatres.2011.04.013.

Latham, Shane J. 2017. "array_split documentation." http://array-split.readthedocs.io/en/latest/.

"Multiprocessing – Process-Based Parallelism." 2017. https://docs.python.org/3/library/multiprocessing.html.

The HDF Group. 1997–1997-NNNN. "Hierarchical Data Format, version 5." http://www.hdfgroup.org/HDF5/.

Van der Walt, Stefan, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. 2014. "Scikit-Image: Image Processing in Python." *PeerJ* 2. PeerJ Inc.: e453. doi:10.7717/peerj.453.

Walt, Stéfan van der, S Chris Colbert, and Gael Varoquaux. 2011. "The Numpy Array: A Structure for Efficient Numerical Computation." *Computing in Science & Engineering*

13 (2). IEEE: 22–30. doi:10.1109/MCSE.2011.37.