The Journal of Open Source Software

root_numpy: The interface between ROOT and NumPy

Edmund Noel Dawe[1], **Piti Ongmongkolkul**[2], **and Giordon Stark**[3]

1 The University of Melbourne **2** Mahidol University International College **3** The University of Chicago

DOI: 10.21105/joss.00307

Software

- Review
- Repository
- Archive

Summary

root_numpy interfaces NumPy (Walt, Colbert, and Varoquaux 2011) with CERN's ROOT (Antcheva et al. 2009) software framework, providing the ability to analyse ROOT data within the broad ecosystem of scientific Python packages.

At its core are functions for converting between ROOT `TTree`s and structured NumPy arrays. root_numpy can convert `TTree` branches (columns) of fundamental types and strings, as well as variable-length and fixed-length multidimensional arrays and (nested) `std::vector`s. root_numpy can also create columns in the output NumPy array from mathematical expressions like ROOT's `TTree::Draw()`. root_numpy's internals are written in Cython (Behnel et al. 2011), installed as compiled C++ extensions, and can handle data with comparable speed to ROOT as shown in the figure below. root_numpy can also convert between ROOT histograms and NumPy arrays, and sample or evaluate ROOT functions as NumPy arrays.

root_numpy interfaces NumPy with TMVA (Speckmayer et al. 2010), ROOT's machine learning toolkit, but naturally allows ROOT users to take advantage of scikit-learn (Pedregosa et al. 2011) and TensorFlow (Abadi et al. 2015).

References

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, et al. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." http://tensorflow.org/.

Antcheva, Ilka, Maarten Ballintijn, Bertrand Bellenot, Marek Biskup, Rene Brun, Nenad Buncic, Philippe Canal, et al. 2009. "ROOT - a C++ Framework for Petabyte Data

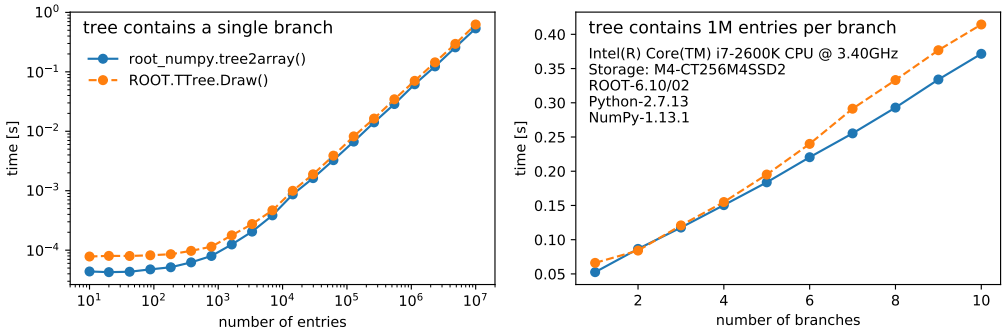

Figure 1: Benchmarking root_numpy's `tree2array()` function against ROOT's `TTree::Draw()`

Storage, Statistical Analysis and Visualization." *Computer Physics Communications* 180 (12): 2499–2512. doi:10.1016/j.cpc.2009.08.005.

Behnel, Stefan, Robert Bradshaw, Craig Citro, Lisandro Dalcin, Dag Sverre Seljebotn, and Kurt Smith. 2011. "Cython: The Best of Both Worlds." *Computing in Science and Engineering* 13: 31–39. doi:10.1109/MCSE.2010.118.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30. http://scikit-learn.org.

Speckmayer, P., A. Hocker, J. Stelzer, and H. Voss. 2010. "The Toolkit for Multivariate Data Analysis, Tmva 4." *J. Phys. Conf. Ser.* 219: 032057. doi:10.1088/1742-6596/219/3/032057.

Walt, Stéfan van der, S. Chris Colbert, and Gaël Varoquaux. 2011. "The Numpy Array: A Structure for Efficient Numerical Computation." *Computing in Science & Engineering* 13: 22–30. doi:10.1109/MCSE.2011.37.