The Journal of Open Source Software

TidyTensor: Utilities for multidimensional arrays as named hierarchical structures

Shawn T. O'Neil[1]

1 Center for Health Artificial Intelligence, University of Colorado Anschutz Medical Campus

DOI: 10.21105/joss.03543

Software

- Review
- Repository
- Archive

Editor: Bita Hasheminezhad

Reviewers:

- @nhejazi
- @omid-s

Submitted: 13 July 2021
Published: 08 October 2021

Summary

Deep learning applications commonly employ the use of *tensors* (multidimensional arrays) for data storage and manipulation. In addition to being space efficient tensors provide a common framework for, the engine of deep neural network training (Baydin et al., 2018). TidyTensor is an R package for inspecting and manipulating tensors. It provides an improved `print()` function for summarizing structure, named tensors, conversion to data frames, and high-level manipulation functions. Designed to complement `keras` package (or other packages that utilize native R arrays), functionality is layered on top of base R types.

Statement of need

While R supports multidimensional arrays natively, there are important differences between these types and tensors as typically used in deep learning applications. First, R assumes array data are organized in a column-major order as opposed to the row-major order typically used in Python, rendering the default `print()` and other inspection methods far less useful. TidyTensor additionally provides support for "named" tensors, allowing researchers to use semantically-relevant names for working with and manipulating tensors via a familiar tidy-friendly interface. While other packages exist for named arrays in R, TidyTensor leverages a hierarchical interpretation of tensor data that makes it easy to manipulate and investigate tensor data of various kinds. Early versions of TidyTensor supported a 3-week researcher-oriented workshop "Deep Learning for Life Scientists" at Oregon State University's Center for Genome Research and Biocomputing, and these features supported graduate students and faculty new to the field in applying deep learning techniques to datasets from their own work.

Related Work

Other packages implement named tensors, though without the "tidy" features of TidyTensor or the simplification of asserting nesting/set-of-sets semantics. In Python, the xarray package provides named tensors and a wide variety of manipulation and related features (Hoyer & Hamman, 2017), while in R the garray package provides features similar to TidyTensor (Huang, 2018), including implementing names in the same way allowing for easy conversion to and from TidyTensors. The stars R package also provides named arrays with a focus on spatiotemporal data (Pebesma, 2021). TidyTensor utilizes abind internally for array manipulation (Plate & Heiberger, 2016).

TidyTensor was designed primarily as a companion to the keras and tensorflow ports for R which utilize native R types (Allaire & Chollet, 2021; Allaire & Tang, 2021). By contrast, the recently released torch port for R does not use native R types for tensors (Falbel & Luraschi, 2021); using TidyTensor with this package would require costly conversion to-and-from Torch-native types.

References

Allaire, J., & Chollet, F. (2021). *keras: R Interface to 'Keras'*. https://CRAN.R-project.org/package=keras

Allaire, J., & Tang, Y. (2021). *tensorflow: R Interface to 'TensorFlow'*. https://CRAN.R-project.org/package=tensorflow

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic Differentiation in Machine Learning: A Survey. *Journal of Machine Learning Research*, *18*. http://jmlr.org/papers/v18/17-468.html

Falbel, D., & Luraschi, J. (2021). *torch: Tensors and Neural Networks with 'GPU' Acceleration*. https://CRAN.R-project.org/package=torch

Hoyer, S., & Hamman, J. (2017). xarray: N-D Labeled Arrays and Datasets in Python. *Journal of Open Research Software*, *5*(1). https://doi.org/10.5334/jors.148

Huang, Q. (2018). *garray: Generalized Array Arithmetic for Ragged Arrays with Named Margins*. https://CRAN.R-project.org/package=garray

Pebesma, E. (2021). *stars: Spatiotemporal Arrays, Raster and Vector Data Cubes*. https://CRAN.R-project.org/package=stars

Plate, T., & Heiberger, R. (2016). *abind: Combine Multidimensional Arrays*. https://CRAN.R-project.org/package=abind