The Journal of Open Source Software

dml: Distance Metric Learning in R

Yuan Tang[1], **Tao Gao**[1], **and Nan Xiao**[2]

1 Ant Financial **2** Seven Bridges

DOI: 10.21105/joss.01032

Software

- Review
- Repository
- Archive

Submitted: 18 October 2018
Published: 19 October 2018

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC-BY).

Summary

Distance metric is widely used in the machine learning literature. We used to choose a distance metric according to a priori (e.g. Euclidean Distance , L1 Distance, etc.) or according to the result of cross validation within small class of functions (e.g. choosing order of polynomial for a kernel). Actually, with priori knowledge of the data, we could learn a more suitable distance metric with (semi-)supervised distance metric learning techniques. `dml` (Y. Tang, Gao, & Xiao, 2015) is such an R package aims to implement the state-of-the-art algorithms for (semi-)supervised distance metric learning.

The `dml` package provides native R implementations of state-of-the-art algorithms for *Distance Metric Learning*, including both global and local methods such as *Relevant Component Analysis*, *Discriminative Component Analysis*, and *Local Fisher Discriminant Analysis* (a list of all the implemented algorithms can be found in the `dml` package reference manual. These methods are widely applied in feature extraction, dimensionality reduction, clustering, information retrieval, and computer vision problems.

Additionally, implementations for the variants of the methods are also available in `dml` package. For example, it builds on top of the `lfda` (Y. Tang, 2017; Tang, Yuan, Li, & Wenxuan, 2016) package to provide access to *Local Fisher Discriminant Analysis* family of methods, which includes *Local Fisher Discriminant Analysis*, *Kernel Local Fisher Discriminant Analysis*, and *Semi-supervised Local Fisher Discriminant Analysis*. To make the results of these methods easy for users to interprete and analyze, both static and interactive visualizations for the results are available through the `ggfortify` (Horikoshi & Tang, 2018; Y. Tang, Horikoshi, & Li, 2016) and the `autoplotly` (Y. Tang, 2018a, 2018b) package respectively.

References

Horikoshi, M., & Tang, Y. (2018). *Ggfortify: Data visualization tools for statistical analysis results*. Retrieved from http://CRAN.R-project.org/package=ggfortify

Tang, Y. (2017). *Local fisher discriminant analysis*. Retrieved from https://CRAN.R-project.org/package=lfda

Tang, Y. (2018a). *Autoplotly: An r package for automatic generation of interactive visualizations for statistical results*. Retrieved from https://CRAN.R-project.org/package=autoplotly

Tang, Y. (2018b). Autoplotly: An r package for automatic generation of interactive visualizations for statistical results. *Journal of Open Source Software*, *3*(24), 657. Retrieved from https://doi.org/10.21105/joss.00657

Tang, Y., Gao, T., & Xiao, N. (2015). *Dml: Distance metric learning in r*. Retrieved from https://CRAN.R-project.org/package=dml

Tang, Y., Horikoshi, M., & Li, W. (2016). Ggfortify: Unified interface to visualize statistical result of popular r packages. *The R Journal*, *8*(2), 478–489. Retrieved from https://journal.r-project.org/

Tang, Yuan, Li, & Wenxuan. (2016). Lfda: An r package for local fisher discriminant analysis and visualization. *arXiv preprint arXiv:1612.09219*. Retrieved from https://arxiv.org/abs/1612.09219