The Journal of Open Source Software

COINSTAC: Collaborative Informatics and Neuroimaging Suite Toolkit for Anonymous Computation

Harshvardhan Gazula[1], **Ross Kelly**[1], **Eric Verner**[1], **Anand D. Sarwate**[1], **Sergey M. Plis**[1], and **Vince D. Calhoun**[1]

1 Center for Translational Research in Neuroimaging and Data Science, Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

DOI: 10.21105/joss.01847

Software

- Review
- Repository
- Archive

Submitted: 29 October 2019
Published: 30 October 2019

Summary

Central to the field of neuroimaging is the development of techniques for making sense of complex brain data. However, rapid technological advancements are pushing the spatial and temporal resolution of imaging in different modalities to a never before seen level leading to large datasets which cannot be worked out in a traditional desktop computing paradigm. This has led to a paradigm shift in scientific research increasing the emphasis on collaborative data-sharing. However, current approaches to data-sharing such as negotiating multiple data sharing agreements, can be cumbersome. In addition, there are also significant data transfer, organization and computational challenges. The bottomline being collaborative group research requires a great deal of coordination. Human and business factors can hamper research from happening at a pace that we are able to handle, maybe even forbidding group research to occur at all.

Software

`COINSTAC` (Plis et al., 2016) is a web-based framework titled Collaborative Informatics and Neuroimaging Suite Toolkit for Anonymous Computation that addresses the aforementioned issues. It provides a platform to analyze data stored locally across multiple organizations without the need for pooling the data at any point during the analysis. It is intended to be an ultimate one-stop shop by which researchers can build statistical (Ming et al., 2017) or machine learning models (Gazula et al., 2018) collaboratively in a decentralized fashion. This framework implements a message passing infrastructure that allows large scale analysis of decentralized data with results on par with those that would have been obtained if the data were in one place. Since, there is no pooling of data it also preserves the privacy of individual datasets. We also offer differentially private algorithms for enhanced protection. Computations can be local or decentralized and are deployed using a containerized model. we also offer a simulation environment for algorithm developers to build COINSTAC computations. These computations can then be made available within the COINSTAC platform.

Features

`COINSTAC` removes the barriers to collaborative analysis by:

1. decentralizing analyses and computation

- Each user performs analyses/pipelines/etc all on their own computers. bits and pieces of each users' output may be sent to a central compute node. Over a dozen local and decentralized computations/algorithms have been developed already with more coming.
- A central compute node performs a complimentary component of the group analysis by coordinating between the various data nodes participating in a consortium. This node may trigger adjusted computations on users' machines, generally in effort to improve a model.

2. not synchronizing full datasets. instead, synchronizing only aggregate level analysis metrics (e.g. the gradients of a machine learning algorithm)

- As previously discussed, central compute nodes aggregate these metrics, and attempt to draw conclusions from the contributor swarm
- Because machine learning algorithms can be designed to model outcomes via artifacts of your analysis Pipelines, we keep your data safely and conveniently on your own machine, untouched.

3. applying differential privacy strategies to further enhance anonymization of private data, whilst still permitting collaboration.

Acknowledgements

This work was funded by the National Institutes of Health (grant numbers: R01EB005846, 1R01DA040487, P20GM103472/5P20RR021938) and the National Science Foundation (grant numbers: 1539067 and 1631819). In addition, the authors would like to acknowledge the efforts of many unnamed personnel over the years who contributed to the development of `COINSTAC`.

References

Gazula, H., Baker, B., Damaraju, E., Plis, S. M., Panta, S. R., Silva, R. F., & Calhoun, V. D. (2018). Decentralized analysis of brain imaging data: Voxel-based morphometry and dynamic functional network connectivity. *Frontiers in neuroinformatics*, *12*, 55.

Ming, J., Verner, E., Sarwate, A., Kelly, R., Reed, C., Kahleck, T., Silva, R., et al. (2017). COINSTAC: Decentralizing the future of brain imaging analysis. *F1000Research*, *6*.

Plis, S. M., Sarwate, A. D., Wood, D., Dieringer, C., Landis, D., Reed, C., Panta, S. R., et al. (2016). COINSTAC: A privacy enabled model and prototype for leveraging and processing decentralized brain imaging data. *Frontiers in neuroscience*, *10*, 365.