HOI: A Python toolbox for high-performance estimation of Higher-Order Interactions from multivariate data

Matteo Neri [1*], **Dishie Vinchhi** [3*], **Christian Ferreyra** [1,5], **Thomas Robiglio** [4], **Onur Ates** [1], **Marlis Ontivero-Ortega** [2], **Andrea Brovelli** [1], **Daniele Marinazzo** [2*], **and Etienne Combrisson** [1*]

1 Institut de Neurosciences de la Timone, Aix Marseille Université, UMR 7289 CNRS, 13005, Marseille, France **2** University of Ghent, Ghent, Belgium **3** Veermata Jijabai Technological Institute, Mumbai **4** Department of Network and Data Science, Central European University, Vienna, Austria **5** Laboratoire d'Informatique et des Systèmes, Aix Marseille Université, UMR 7020 CNRS, Marseille, France ***** These authors contributed equally.

DOI: 10.21105/joss.07360

Software

- Review
- Repository
- Archive

Editor: Mehmet Hakan Satman

Reviewers:

- @pitmonticone
- @ClaudMor

Submitted: 24 September 2024
Published: 12 November 2024

Summary

`HOI` (Higher-Order Interactions) is a Python toolbox to measure higher-order information theoretic metrics from multivariate data. Higher-order interactions refer to interactions that go beyond pairwise connections between nodes in a network (Battiston et al., 2021; Baudot et al., 2019; Gatica et al., 2021; Herzog et al., 2022; Luppi et al., 2024; Rosas et al., 2019). The `HOI` toolbox provides easy-to-use information theoretical metrics to estimate pairwise and higher-order information from multivariate data. The toolbox contains cutting-edge methods, along with core entropy and mutual information functions, which serve as building blocks for all metrics. In this way, `HOI` is accessible both to scientists with basic Python knowledge using pre-implemented functions and to experts who wish to develop new metrics on top of the core functions. Moreover, the toolbox supports computation on CPUs and GPUs. Finally, `HOI` provides tools for visualizing and presenting results to simplify the interpretation and analysis of the outputs.

Statement of need

Recent research studying higher-order interactions with information theoretic measures provides new angles and valuable insights in different fields, such as neuroscience (Baudot et al., 2019; Combrisson et al., 2024; Gatica et al., 2021; Herzog et al., 2022; Luppi et al., 2022), music (Rosas et al., 2019), economics (Scagliarini et al., 2023), and psychology (Marinazzo et al., 2022). Information theory allows investigating higher-order interactions using a rich set of metrics that provide interpretable values of the statistical interdependency among multivariate data (Barrett, 2015; Mediano et al., 2021; Rosas et al., 2019; Scagliarini et al., 2023; Williams & Beer, 2010).

Despite the relevance of studying higher-order interactions across various fields, there is currently no toolkit that compiles the latest approaches and offers user-friendly functions for calculating higher-order information metrics. Computing higher-order information presents two main challenges. First, these metrics rely on entropy and mutual information, whose estimation must be adapted to different types of data (Czyż et al., 2024; Madukaife & Phuc, 2024). Second, the computational complexity increases exponentially as the number of variables and interaction orders grows. For example, a dataset with 100 variables, has approximately

1.6×10^5 possible triplets, 4×10^6 quadruplets, and 7×10^7 quintuplets. Therefore, an efficient implementation, scalable on modern hardware is required.

Related packages

Several toolboxes have implemented a few HOI metrics like `infotopo` (Baudot et al., 2019), `infotheory` (Candadai & Izquierdo, 2020) in C++, `DIT` (James et al., 2018), `IDTxl` (Wollstadt et al., 2019) and `pyphi` (Mayner et al., 2018), in Python. However, `HOI` is the only pure Python toolbox specialized in the study of higher-order interactions offering functions to estimate with an optimal computational cost a wide range of metrics as the O-information (Rosas et al., 2019), the topological information (Baudot et al., 2019) and the redundancy-synergy index (Timme & Lapish, 2018). Moreover, `HOI` allows to handle Gaussian, non-Gaussian, and discrete data using different state-of-the-art estimators (Czyż et al., 2024; Madukaife & Phuc, 2024). `HOI` also distinguishes itself from other toolboxes by leveraging `Jax`, a library optimized for fast and efficient linear algebra operations on both CPU, GPU and TPU. Taken together, `HOI` combines efficient implementations of current methods and is adaptable enough to host future metrics, facilitating comparisons between different approaches and promoting collaboration across various disciplines.

Acknowledgements

We acknowledge the support from Google via the Summer of Code program via the International Neuroinformatics Coordination Facility initiative. M.N. have received funding from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference : ANR-16-CONV000X / ANR-17-EURE-0029) and from Excellence Initiative of AixMarseille University - A*MIDEX (AMX-19-IET-004). A.B. and E.C were supported by the PRC project "CausaL" (ANR-18-CE28-0016) and received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). A.B. was supported by A*Midex Foundation of Aix-Marseille University project "Hinteract" (AMX-22-RE-AB-071). The "Center de Calcul Intensif of the Aix-Marseille University (CCIAM)" is acknowledged for high-performance computing resources. A.B, E.C and D.M were supported by EU's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 101147319 (EBRAINS 2.0 Project). C.F. was supported by the French National Research Agency (ANR-21-CE37-0027). We thank Giovanni Petri for fruitful suggestions and discussions.

References

Barrett, A. B. (2015). Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E*, *91*, 052802. https://doi.org/10.1103/PhysRevE.91.052802

Battiston, F., Amico, E., Barrat, A., Bianconi, G., Ferraz de Arruda, G., Franceschiello, B., Iacopini, I., Kéfi, S., Latora, V., Moreno, Y., & others. (2021). The physics of higher-order interactions in complex systems. *Nature Physics*, *17*(10), 1093–1098. https://doi.org/10.1038/s41567-021-01371-4

Baudot, P., Tapia, M., Bennequin, D., & Goaillard, J.-M. (2019). Topological information data analysis. *Entropy. An International and Interdisciplinary Journal of Entropy and Information Studies*, *21*(9). https://doi.org/10.3390/e21090869

Candadai, M., & Izquierdo, E. J. (2020). Infotheory: A C++/Python package for multivariate information theoretic analysis. *Journal of Open Source Software*, *5*(47), 1609. https://doi.org/10.21105/joss.01609

Combrisson, E., Basanisi, R., Gueguen, M. C., Rheims, S., Kahane, P., Bastin, J., & Brovelli, A. (2024). Neural interactions in the human frontal cortex dissociate reward and punishment learning. *eLife*, *12*, RP92938. https://doi.org/10.7554/eLife.92938

Czyż, P., Grabowski, F., Vogt, J., Beerenwinkel, N., & Marx, A. (2024). Beyond normal: On the evaluation of mutual information estimators. *Advances in Neural Information Processing Systems*, *36*.

Gatica, M., Cofré, R., Mediano, P. A., Rosas, F. E., Orio, P., Diez, I., Swinnen, S. P., & Cortes, J. M. (2021). High-order interdependencies in the aging brain. *Brain Connectivity*, *11*(9), 734–744. https://doi.org/10.1089/brain.2020.0982

Herzog, R., Rosas, F. E., Whelan, R., Fittipaldi, S., Santamaria-Garcia, H., Cruzat, J., Birba, A., Moguilner, S., Tagliazucchi, E., Prado, P., & others. (2022). Genuine high-order interactions in brain networks and neurodegeneration. *Neurobiology of Disease*, *175*, 105918. https://doi.org/10.1016/j.nbd.2022.105918

James, R. G., Ellison, C. J., & Crutchfield, J. P. (2018). "Dit": A Python package for discrete information theory. *Journal of Open Source Software*, *3*(25), 738. https://doi.org/10.21105/joss.00738

Luppi, A. I., Mediano, P. A., Rosas, F. E., Holland, N., Fryer, T. D., O'Brien, J. T., Rowe, J. B., Menon, D. K., Bor, D., & Stamatakis, E. A. (2022). A synergistic core for human brain evolution and cognition. *Nature Neuroscience*, *25*(6), 771–782. https://doi.org/10.1038/s41593-022-01070-0

Luppi, A. I., Rosas, F. E., Mediano, P. A., Menon, D. K., & Stamatakis, E. A. (2024). Information decomposition and the informational architecture of the brain. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2023.11.005

Madukaife, M. S., & Phuc, H. D. (2024). Estimation of Shannon differential entropy: An extensive comparative review. *arXiv Preprint arXiv:2406.19432*.

Marinazzo, D., Van Roozendaal, J., Rosas, F. E., Stella, M., Comolatti, R., Colenbier, N., Stramaglia, S., & Rosseel, Y. (2022). An information-theoretic approach to hypergraph psychometrics. *arXiv Preprint arXiv:2205.01035*.

Mayner, W. G., Marshall, W., Albantakis, L., Findlay, G., Marchman, R., & Tononi, G. (2018). PyPhi: A toolbox for integrated information theory. *PLoS Computational Biology*, *14*(7), e1006343. https://doi.org/10.1371/journal.pcbi.1006343

Mediano, P. A., Rosas, F. E., Luppi, A. I., Carhart-Harris, R. L., Bor, D., Seth, A. K., & Barrett, A. B. (2021). Towards an extended taxonomy of information dynamics via integrated information decomposition. *arXiv Preprint arXiv:2109.13186*.

Rosas, F. E., Mediano, P. A. M., Gastpar, M., & Jensen, H. J. (2019). Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Physical Review E*, *100*(3), 032305. https://doi.org/10.1103/PhysRevE.100.032305

Scagliarini, T., Nuzzi, D., Antonacci, Y., Faes, L., Rosas, F. E., Marinazzo, D., & Stramaglia, S. (2023). Gradients of O-information: Low-order descriptors of high-order dependencies. *Phys. Rev. Res.*, *5*, 013025. https://doi.org/10.1103/PhysRevResearch.5.013025

Timme, N. M., & Lapish, C. (2018). A tutorial for information theory in neuroscience. *eNeuro*, *5*(3). https://doi.org/10.1523/ENEURO.0052-18.2018

Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information. *arXiv Preprint arXiv:1004.2515*.

Wollstadt, P., Lizier, J. T., Vicente, R., Finn, C., Martinez-Zarzuela, M., Mediano, P., Novelli, L., & Wibral, M. (2019). IDTxl: The information dynamics toolkit xl: A Python package for the efficient analysis of multivariate information dynamics in networks. *Journal of Open*

Source Software, *4*(34), 1081. https://doi.org/10.21105/joss.01081