



Editorial: Applications of Machine Learning to Evolutionary Ecology Data

Juliano Morimoto^{1*}, Aurore Ponchon¹, Georgy Sofronov² and Justin Travis¹

¹ School of Biological Sciences, University of Aberdeen, Aberdeen, United Kingdom, ² Department of Mathematics and Statistics, Macquarie University, Sydney, NSW, Australia

Keywords: machine learning, artificial intelligence, ecological theories, innovation, niche adaptation

Editorial on the Research Topic

Applications of Machine Learning to Evolutionary Ecology Data

“Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom” – Stoll (1995)

Machine Learning (ML) has become an increasingly popular tool in a large variety of fields due to an explosion of ways in which large-scale datasets are collected. ML has been used both as the means through which large data is generated (e.g., image recognition and classification) as well as the tools through which information and knowledge are acquired from large data through complex processing of high-dimensional data (James et al., 2013; Kroese et al., 2019). Thus, ML holds the potential to revolutionize the way in which humanity makes sense of the vast amount of data generated every day. This Research Topic includes a collection of articles which give an overview of applications of ML methods to evolutionary ecology data. In particular, the articles in Research Topic provide innovative insights into how ML can be used to facilitate taxonomic identification and biodiversity survey, find new relationships between morphology and fitness, digital imaging processing, and quantifying ecological network complexity (a key parameter to predict extinction potential of ecological networks).

Gerovichev et al. develop a deep learning model to identify three flying insects. Using a clever combination of a complex deep learning algorithm with simple sticky traps, the authors propose a new high-throughput method to monitor insect populations. This application of ML has the potential to revolutionize the way in which insect surveys are conducted, addressing the major bottleneck in biodiversity assessments of insects (i.e., high-throughput identification). Given the growing concerns of insect population declines (Fox et al., 2019; Didham et al., 2020), this high-throughput framework can enable large-scale insect biodiversity surveys, with ultimate impact to biodiversity conservation.

Lürig et al. promote the use of computer vision, the automated extraction and processing of information contained in digital images, as a promising powerful tool to collect and analyze complex multidimensional phenotype data. After introducing the history and recent development of computer vision methods, the authors propose some general guidelines for beginners willing to dive into computer vision, notably before and after generating digital images. The authors also provided examples of diverse applications of the method in an ecological and evolutionary context. The authors concluded that computer vision methods should provide higher throughput, in a more reproducible way and should facilitate the study of complex phenomics data.

Rabinovich proposes an application of machine learning as a regression and classification tool to highlight relationships between some morphological traits, demographic parameters, environmental factors, and fitness components in kissing bugs, a group of insects vectors of

OPEN ACCESS

Edited and reviewed by:

Elise Huchard,
UMR5554 Institut des Sciences de
l'Evolution de Montpellier (ISEM),
France

*Correspondence:

Juliano Morimoto
juliano.morimoto@abdn.ac.uk

Specialty section:

This article was submitted to
Behavioral and Evolutionary Ecology,
a section of the journal
Frontiers in Ecology and Evolution

Received: 18 October 2021

Accepted: 10 November 2021

Published: 30 November 2021

Citation:

Morimoto J, Ponchon A, Sofronov G
and Travis J (2021) Editorial:
Applications of Machine Learning to
Evolutionary Ecology Data.
Front. Ecol. Evol. 9:797319.
doi: 10.3389/fevo.2021.797319

the parasite *Trypanosoma cruzi*, agent of Chagas disease. First, he reviews synthetically all the accumulated empirical evidence of relationships between various traits of kissing bugs and three specific variables: temperature, feeding frequency, and blood sources. Then, from his review, he extracts 11 variables to analyse their effects on kissing bug fitness with the Random Forest approach and compare it with the Ordinary Least Square Regression method. He showed that the Random Forest approach is a more powerful predictive tool compared to more classical statistical tools but the results reveal fewer relationships than expected. This example of application of machine learning opens new avenues for a better understanding on how multiple inter-related morphological, demographic, and environmental factors affect species fitness.

Studying ecological networks requires the use of different measures to describe their complexity and understand their structure and functioning. Yet, network complexity has various definitions which measures describe the behavior of the system but there are not always appropriate to predict its stability. Strydom et al. present relative rank deficiency and Singular Value Decomposition (SVD) entropy as measures of external and internal physical complexity of ecological networks, which provides insights onto the amount of information required to encode the system and how much signal this information contains. In particular, the authors applied machine learning techniques of simulation as well as statistical and mathematical models to estimate two measures of complexity: one external to the system, the rank of the matrix, which is based on the number of unique interrelationships between modules composing the network, and one internal, the SDV entropy measure, which reveals the importance of each rank in reconstructing the network. The authors showed that while the rank of the matrix

holds little information, SVD entropy is an intuitive, robust, non-structural approach to determining the physical complexity of ecological networks. The authors also showed that complex networks are not more robust to extinction than simpler ones and that larger networks are not necessarily more complex than smaller ones. The use of machine learning allowed to manage and analyse large datasets of species interaction networks and work with multidimensional data. Overall, the use of SDV entropy is proposed as a new starting point to define the complexity of ecological networks, which will further help describe and understand the relationship between ecological network complexity and stability.

ML is revolutionizing our understanding of natural patterns, by enabling complex analysis of large-scale datasets. The articles of this Research Topic form a collection of highly innovative ways that ML has been implemented to advance the field of Ecology and Evolution. Our hope is that the Research Topic can serve as a useful resource to the academic community for further developments in the field of ecoinformatics, which enable more and bigger discoveries to aid our understanding of how living organisms interact with each other and their environment.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We acknowledge the authors that kindly supported our initiative and shared their work in this Research Topic.

REFERENCES

- Didham, R. K., Basset, Y., Collins, C. M., Leather, S. R., Littlewood, N. A., Menz, M. H. M., et al. (2020). Interpreting insect declines: seven challenges and a way forward. *Insect Conserv. Divers.* 13, 103–114. doi: 10.1111/icad.12408
- Fox, R., Harrower, C. A., Bell, J. R., Shortall, C. R., Middlebrook, I., and Wilson, R. J. (2019). Insect population trends and the iucn red list process. *J. Insect Conserv.* 23, 269–278. doi: 10.1007/s10841-018-0117-1
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York, NY: Springer.
- Kroese, D. P., Botev, Z. I., Taimre, T., and Vaisman, R. (2019). *Data Science and Machine Learning: Mathematical and Statistical Methods*. Boca Raton: Chapman and Hall/CRC.
- Stoll, C. (1995). *Silicon Snake Oil: Second Thoughts on the Information Highway*. Garden City, NY: Anchor Books/Doubleday & Co.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Morimoto, Ponchon, Sofronov and Travis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.