Eur. Phys. J. Spec. Top. (2021) 230:2173–2175
https://doi.org/10.1140/epjs/s11734-021-00275-x

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Editorial

Measure or infer? Role of modeling and machine learning in modern astronomy

Snehanshu Saha[1] and Nithin Nagaraj[2,a]

[1] CS&IS and Anuradha and Prashanth Palakurthi Centre for Artificial Intelligence Research (APPCAIR), BITS Pilani, K K Birla Goa Campus, Sancoale, India
[2] National Institute of Advanced Studies, Indian Institute of Science Campus, Bengaluru, India

Published online 7 September 2021

Theory of machine learning, deep learning in particular has been witnessing an implosion lately in deciphering the "black-box approaches". Optimizing deep neural networks is largely thought to be an empirical process, requiring manual tuning of several parameters. Drawing insights in to these parameters gained much attention lately. Data is at the heart of this. Astronomy is a fascinating case study as it had embraced big data embellished by many sky-surveys. The variety and complexity of the data sets at different wavelengths, cadences, etc. imply that modeling, computational intelligence methods and machine learning are imperative to understand astronomy. The importance of data driven discovery in astronomy has given birth to an exciting new field known as Astroinformatics. This inter-disciplinary study brings together machine learning theorists, astronomers, mathematicians and computer scientists underpinning the importance of machine learning algorithms and data analytic techniques.

As Editors of the Special issue, our goal was to bring about a scholarly collection of articles including original research and well-thought out, insightful reviews at the intersection of Machine Learning and Astronomy that will set a unique ground as an amalgamation of the diverse ideas and techniques while staying true to the baseline. The special issue managed to assemble stellar collection of ideas that discuss new developments in modeling, machine learning, design of complex computer experiments and data analytic techniques which can be used in areas beyond astronomical data analysis. Given the horizontal nature of the issue, we believe that we have been able to disseminate methods that are largely area-agnostic but currently of interest to the broad community of Astronomy. While the goal was to publish an excellent set of articles, the larger objective of the issue was to generate interest in Machine Learning and Statistical techniques among the Astronomy community. The special issue will provide skilful introductions, reproducible codebases and insightful reviews for the community. We are pleased to state that the issue covers topics from the following areas:

- Exoplanets (discovery, machine classification etc.)
- Unsupervised, semi-supervised, and supervised representation learning
- Metric learning and kernel learning
- Deep learning in astronomy
- MCMC on big data
- Statistical machine learning
- Bayesian methods in astronomy
- Meta-heuristic and evolutionary clustering methods and applications in astronomy
- Optimization methods
- Swarm intelligence
- Multi-objective optimization
- Information-theoretic methods in life-like systems

Random forest has recently been applied in a variety of astronomical problems—owing to the ease with which it can handle high-dimensional data without data normalization, Md. Mahumdunnobe et al. [1] employ supervised random forest to estimate membership of a large sample of Gaia DR2 data for nine open clusters. Their method makes it possible to find sub-structure in both position and velocity space of the clusters. They report a precision of around 90% indicating the efficacy of random forest for star cluster studies.

Alexei Nekrasov et al. [2] study the fascinating subject of interstellar extinction. They provide an analytical tri-dimensional model of interstellar extinction.

Life on other planets has always fascinated the human mind. Characterizing habitability of the billions of planets in our galaxy alone is a very challenging task. Exoplanetary habitability is a highly inter-disciplinary subject interfacing astrophysics, astrobiology, planetary science and other allied sciences. Margarita Safonova et al. [3] review existing metrics of habitability and new classification schemes of extrasolar planets—including

[a] e-mail: nithin@nias.res.in (corresponding author)

those that apply machine learning (ML) and artificial neural networks (ANN) with a variety of activation functions.

Archana Mathur et al. [4] delve further into this challenging problem of classification of exoplanets using ML. They propose a paradigm to automate the task of exoplanet classification by performing a detailed analysis of the structure of the Puerto Rico's Planetary Habitability Laboratory's Exoplanet Catalog dataset (PHL-EC). The study focuses on novel neural network architectures for this classification task, and the development of a general data methodology, and a set of best practices that can be applied for exploratory data analysis experiments.

Joshua Lukemire et al. [5] address the problem of time-consuming sophisticated computer simulations used by scientists to test complex systems by exploring computationally cheap surrogates to facilitate analysis and optimization of the underlying system. Their study discusses design criteria that allow astronomy researchers to extract the maximum benefit offered by Gaussian Process surrogate modeling (for example—identification of optimal Latin hypercube designs).

The bi-objective optimization problem arising out of habitability score computation of exoplanets is given an innovative treatment by Luckyson Khaidem et al. [6] by drawing an analogy and solving the constrained optimization problem in production economics. A novel multi-objective implementation of Quantum Particle Swarm Optimization (QPSO) is described. A game theoretic interpretation of the components of the bi-objective framework is also presented.

This special issue also presents four excellent review articles interfacing machine learning, deep learning, optimization and astronomy. Written in tutorial style, they assist the beginner to get a clear grasp of important concepts, ideas and algorithms in these interdisciplinary fields.

Saroj Meher and Ganapti Panda [7] discuss learning aspects of ML, Deep Learning (DL) including selecting, extracting and pre-processing the input data, advantages and disadvantages of using these methods on different related tasks of astronomy using massive amount of domain data from various diverse sources.

Multi-objective evolutionary algorithms have gained popularity since 1980s with significant developments in the last two decades. Owing to their flexibility and adaptable nature, they find applications in a wide range of disciplines such as—astronomy and astrophysics, bioinformatics, image processing, natural language processing etc. Naveen Saini and Sriparna Saha [8] survey recently developed multi-objective optimization algorithms. Potential future research and applications of these are also discussed.

Sravan Danda et al. [9] provide an in-depth tutorial review consolidating the existing applications of the power watershed optimization framework in the context of image processing. This is primarily aimed at researchers working on image filtering and segmentation algorithms in application areas such as astronomy where images typically have a large number of pixels. PW versions of several popular graph-based approaches to image segmentation and filtering (such as random walker segmentation, isoperimetric partitioning, ratio cut, multi-cut and shortest path edge-preserving filters) are visualized as contrast-agnostic fast approximations to the corresponding methods.

Rahul Yedida and Snehanshu Saha [10] provide a primer on ML for beginners. Though the number of excellent books on this subject is increasing at a rapid pace, this primer aims to provide an accessible and largely conversational (sans jargons) introduction to the subject while not compromising on mathematical details, and provides implementation of most algorithms in Python.

The painstaking efforts by all the expert reviewers ensured that each submission was rigorously evaluated in a timely manner, and every article underwent rounds of revision. It was a pleasure working with the Editorial office, in particular Mrs. Sandrine Karpe and Prof. B. Ananthnaryanan. We are also immensely grateful to helpful suggestions and lengthy discussions with faculty colleagues coming from a diverse portfolio of Scientific interests.

We hope that this volume can enable constructive explorations and collaborative discussions between researchers in the fields of modeling, machine learning and astronomy, thereby stimulating creative interdisciplinary research.

References

1. M. Mahmudunnobe, P. Hasan, M. Raja et al., Membership of stars in open clusters using random forest with Gaia data. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00205-x
2. A. Nekrasov, K. Grishin, D. Kovaleva et al., Approximate analytical description of the high latitude extinction. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00210-0
3. M. Safonova, A. Mathur, S. Basak et al., Quantifying the classification of exoplanets: in search for the right habitability metric. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00211-z
4. S. Basak, A. Mathur, A.J. Theophilus et al., Habitability classification of exoplanets: a machine learning insight. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00203-z
5. J. Lukemire, Q. Xiao, A. Mandal et al., Statistical analysis of complex computer models in astronomy. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00204-y
6. L. Khaidem, S. Saha, S. Kar et al., Expert habitat: a colonization conjecture for exoplanetary habitability via penalized multi-objective optimization-based candidate validation. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00208-8
7. S.K. Meher, G. Panda, Deep learning in astronomy: a tutorial perspective. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00207-9

8. N. Saini, S. Saha, Multi-objective optimization techniques: a survey of the state-of-the-art and applications. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00206-w
9. S. Danda, A. Challa, B.S. Daya Sagar, L. Najman, A tutorial on applications of power watershed optimization to image processing. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00264-0
10. R. Yedida, S. Saha, Beginning with machine learning: a comprehensive primer. Eur. Phys. J. Spec. Top. (2021). https://doi.org/10.1140/epjs/s11734-021-00209-7