S11: Machine Learning, a giant leap towards space discovery in the era of peta and exabyte scale surveys

The members of the organising committee are as follows

Brescia M. (chair, INAF-OACN, Italy)
Chamba N. (chair, Stockholm University, Sweden)
Lazar I. (chair, University of Hertfordshire, UK)
Martin G. (chair, KASI, Korea / Steward Observatory, USA)
Tortora C. (chair, INAF-OACN, Italy)
Borgani S. (INAF-OATS, Italy)
Huertas-Company M. (Observatory of Paris, France)
Jaodand A. (Caltech, USA)
Kaviraj S. (University of Hertfordshire, UK)
Mei S. (Observatory of Paris, France)
Salvato M. (MPE, Germany)
Sarmiento R. (IAC, Spain)
Scaife A. (University of Manchester, UK)
Wenskovitch J. (Virginia Tech, USA)

Foreword

Astronomy and cosmology are among the forefront disciplines in Big Data science, with exponentially growing data volumes, rates and complexity. Data from current and future generation "Big Data" missions and telescopes will reach the peta and exabyte regimes and become the new normal in future astrophysical research. These datasets will require game-changing statistics and provide high-quality data that is essential for tackling key open questions such as the energy and mass content of the Universe or the major physical quantities that drive galaxy formation.

"Machine Learning: a giant leap towards space discovery in the era of peta and exabyte scale surveys" held between 30th July and 1st June 2022 was a Symposium held as part of the European Astronomical Society annual meeting, which took place in Valencia, Spain. The Symposium brought together leading astronomers and data scientists working to apply machine learning techniques to astronomical questions, showcasing the exciting possibilities that machine learning and artificial intelligence offer to the astronomy and cosmology community, particularly in the face of the rapidly growing data volumes, rates and complexity from current and future missions and telescopes.

The Symposium's purpose was to discuss the challenges inherent in the extraordinary volume of these datasets and how to overcome them. In anticipation of Big Data surveys, the community has increasingly invested in developing advanced machine learning (ML) and artificial intelligence (AI) based data-analysis techniques. These techniques provide the level of accuracy and automation required for the efficient exploitation of extremely large datasets and are quickly becoming the default choice for many astronomers. AI can analyse a vast amount of data for diverse scientific aims, from cosmology to stellar physics, extragalactic astronomy, and planetary science.

Exascale computing will provide the capability to tackle challenges in scientific discovery at levels of complexity and performance that were previously unattainable, enabling the full exploitation of multi-messenger Astrophysics and numerical simulations for astrophysics and cosmology. Given the imminent arrival of extremely large surveys, the Symposium was particularly timely, and we hope it enabled researchers to build collaborations that will benefit the astronomical community for many years to come.

We were extremely pleased by the success of the Symposium, which featured 5 distinguished invited talks and over 40 exceptional oral presentations and posters presented over 2 days. The Symposium drew a significant contribution from early-career researchers, adding diversity and fresh perspectives to the discussions. Participants addressed key challenges in the current state-of-the-art for machine learning, including applications in existing datasets, new developments in computer science, and steps towards implementing new tools to exploit large datasets to their fullest potential as the capabilities of instruments and their datasets continue to evolve with the increasing fidelity, cadence and data volume of observations. The range of pertinent topics covered at the Symposium was impressive, spanning cosmology and large-scale structure, galaxy morphology, merger identification, source identification, deblending, time-domain astronomy, and more. The methods and techniques showcased at the Symposium were diverse, featuring generative models, the interface of machine learning with citizen science initiatives and simulations, unsupervised learning, deep learning, and many others.

The conference included altogether five sessions with invaluable presentations from invited speakers and contributed talks. The first session began with an invited talk by Helena Domınguez Sanchez, who gave an excellent overview of applications of deep learning algorithms in the morphological classification of galaxies. The session continued with a number of other talks showcasing methods for characterising and quantifying galaxy structure, which included a discussion of the exploitation of large-scale human labelling from citizen science initiatives and the current limitations and pitfalls of current techniques, especially with respect to rare populations and how this interfaces with the huge scale of unlabelled images that new surveys will produce.

The second session began with an invited talk by Claudio Grillo, who reviewed applications of machine learning to problems related to strong gravitational lensing. This was followed up by talks broadly on the theme of extracting physical insights about astrophysical objects and model fitting and comparison with extremely large datasets.

Next Michelle Ntampaka discussed creating trustworthy ML to interpret upcoming large surveys and enabling physical discovery. In later discussions, we found that the interpretability and trustworthiness of techniques were a common concern in addition to the compatibility and trustworthiness of the training data input into these models. The rest of the session focussed on overcoming the challenge of source blending, which has become increasingly pressing as the increasing depth of new surveys begins to yield huge increases in the number of astrophysical sources that can be detected.

The fourth session began with an invited talk by Francisco Villaescusa Navarro, who summarised efforts to solve a number of outstanding problems in cosmology by combining huge datasets from state-of-the-art numerical simulations with machine learning methods. Other talks also focussed on applications of machine learning to cosmological problems. We learned about efforts to extract and infer the properties of large-scale structure using both generative adversarial network and swarm-intelligence based methods as well as methods for inferring the properties of galaxy dark matter haloes and their relationship with their central galaxies.

In our final session, Kerstin Bunte explained how principles of evolutionary computation, topology, swarm optimisation and probabilistic modelling are being used in the recovery of the geometrical structure of faint filamentary structures in a range of astrophysical contexts. We continued with talks focussing on a number of topics, including the classification of galaxy spectra and galaxy morphology and uncertainty-aware machine learning.

During a final discussion session and throughout the Symposium, we sketched out an overview of the current position of the community as well as perspectives for the future. In addition, we identified a number of high-priority questions that should be explored in the near future: (1) The limitations of the usefulness of machine learning, in particular use cases where efficacy over simpler methods may not be clear. (2) Uncertainties and uncertainty propagation including standardisation of methods for different techniques. (3) Can machine learning give us new insights into current astronomical studies? For example, through novel classification schemes to replace the ones that are defined by humans such as the Hubble sequence for galaxy morphology or BPT for AGN. (4) The possibility of an EAS machine learning working group and at least establishing better communication between machine learning projects within missions including encouraging data providers to provide catalogues that can be more easily compared/understood through more standardised and clear descriptions of the data.

We sincerely thank all participants and speakers for their thought-provoking contributions and especially those members of the local organising committee, venue staff and others whose on-the-ground efforts were essential to keeping things running smoothly and enabling thorough scientific discussions. The papers included in these conference proceedings provide a comprehensive overview of the discussions and highlight the progress made (and in some cases highlighting the limitations) in applying machine learning techniques to astronomical questions. Given the rapid rate of progress and the imminent arrival of a number of new large surveys, we hope that these proceedings will give valuable insight into the opportunities and solutions to new challenges these datasets will bring in the near future and serve as a valuable reference for astronomers, data scientists, and anyone interested in the field of machine learning applied to astrophysics research.