EDITORIAL

Introducing Machine Learning: Science and Technology

O Anatole von Lilienfeld
Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), Department of Chemistry, University of Basel, 4056 Basel, Switzerland
E-mail: anatole.vonlilienfeld@unibas.ch

Abstract
Due to the remarkable progress of ever-growing digitalisation and computing capabilities, data has become increasingly abundant, and machine learning has emerged as a key ingredient in many enabling technologies within modern society. Its potential for pushing the frontiers of science is now also clear and has been demonstrated in various domains extending from novel materials design, quantum physics and the simulation of molecules and chemical systems, to particle physics, medical imaging, space science, climate science and drug discovery. Conceived in close consultation with the community, Machine Learning: Science and Technology has been launched as a unique multidisciplinary, open access journal that will bridge the application of machine learning across the natural sciences with new conceptual advances in machine learning methods as motivated by physical insights.

1. Motivation

On behalf of the editorial board and IOP Publishing, I am pleased to announce the opening of Machine Learning: Science and Technology (MLST) with our first published content:

‘Embedding human heuristics in machine-learning-enabled probe microscopy’, Oliver Gordon, Filipe Junqueira, Filipe, and Philip Moriarty [1]
‘Reinforcement learning for semi-autonomous approximate quantum eigensolver’, Francisco Albarrán-Arriagada, Juan Carlos Retamal, Enrique Solano and Lucas Lamata [2]
‘Repetitive readout enhanced by machine learning’, Genyue Liu, Mo Chen, Yi-Xiang Liu, David Layden, and Paola Cappellaro [3]
‘Regularised atomic body-ordered permutation-invariant polynomials for the construction of interatomic potentials’, Cas van der Oord, Geneviève Dusson, Gábor Csányi, Gabor, and Christoph Ortner [4]
‘Improving the background of gravitational-wave searches for core collapse supernovae: a machine learning approach’, Marco Cavaglia, Sergio Gaudio, Travis Hansen, Kai Staats, Marek Szczepanczyk, and Michele Zanolin [5]
‘Image mapping the temporal evolution of edge characteristics in tokamaks using neural networks’, Vignesh Gopakumar and Debasmita Samaddar [6]

Throughout history we can see evidence of statistical learning being applied to solve problems and aid discovery. Specific applications, often characterised by exceptionally high societal interest paired with little or incomplete understanding, have included efforts to predict outcomes in problems as diverse as they were difficult, e.g. crop management, gold mining, stock trading, or drug discovery. The general appeal lies in the promise to gain new perspectives through discovery of patterns, structure, or quantitative numerical models, exclusively based on correlations encoded in stand-alone data, i.e. even for unknown relationships and dependencies. Nowadays, the ever-increasing data abundance due to ever-decreasing cost of computations has led to machine learning solutions for, among others, selling products online (ads, books, movies, etc), brain-computer interfaces, translation services, or autonomous navigation.

Exploiting implicit structure and correlations in large and complex data sets, unsupervised learning enables meaningful classification and categorisation without heuristic or human bias. Supervised learning, by contrast,
relates input variables (features) to output (label) through regression to training data using generic basis functions. Resulting surrogate models then infer, in the limit of infinite training data density and interpolative regimes, smooth property function values for out-of-sample instances with arbitrary precision. In either case, the conditio sine qua non is the availability of data with sufficient quantity and quality.

Due to rapid technological advancements over the last decades, there is tremendous potential for applying statistical learning to vast experimental data sets routinely generated by modern scientific equipment such as MRIs, x-rays, particle colliders, high-throughput robotics, or space telescopes. While the exact relationships among some of these properties recorded are yet to be discovered, there are many other important problems for which the relevant equations (or approximations thereof) are already known, yet are too difficult to solve analytically. In these cases, using numerical solvers and high-performance computing, extensive synthetic data sets can be generated, rendering also these problems amenable to machine learning. Examples include approximate solutions to Schrödinger’s equation for materials and molecules, or to Navier–Stokes equations for turbulent flow and efficient sampling of complex statistical distributions of very high dimension.

If we consider experimentation, theory, and computation, to represent the chronologically and sequentially emerging pillars of science, each building on the previous pillar(s), it is becoming increasingly clear that a fourth pillar, driven by data and machine learning techniques, is currently emerging. This trend is reflected in the number of special issues devoted solely to machine learning, data science, and artificial intelligence driven contributions that have been organised over the last couple of years by leading scientific domain-specific journals.

It is with this backdrop that we firmly believe that the timing is right to introduce Machine Learning: Science and Technology (MLST) as the world’s first central venue for all data driven machine learning developments relevant to the advancement of the exact sciences. MLST is explicitly not domain specific, but rather serves an interdisciplinary author- and readership at the boundaries of the traditional branches, and through open science principles provides the maximum possible dissemination and reuse of content in the interest of accelerating new scientific discovery. Since physics represents the neutral common denominator where machine learning advances across all areas of science can come together, under the common language of applied mathematics, I am also personally delighted that MLST is being launched by IOP Publishing as a leading society publisher in the field.

2. Scope

MLST is a multidisciplinary journal that bridges the application of machine learning across the sciences with new conceptual advances in machine learning methods as motivated by physical insights. Specifically, articles are expected to (i) make conceptual or methodological advances in machine learning with applications to (or motivated by) scientific problems, or (ii) to advance the state of the art of machine learning-driven applications in the sciences.

Particular areas of scientific application include (but are not limited to):

- Physics and space science;
- Design and discovery of novel materials and molecules;
- Materials characterisation techniques;
- Simulation of materials, chemical processes and biological systems;
- Atomistic and coarse-grained simulation;
- Quantum computing;
- Biology, medicine and biomedical imaging;
- Geoscience (including natural disaster prediction) and climatology;
- Simulation methods and high-performance computing.

New conceptual advances in machine learning methods (such as explainability, causality and robustness) include (but are not limited to):

- New learning algorithms;
- Deep learning architectures;
- Kernel methods;
• Probabilistic and Bayesian methods;
• Generative methods;
• Reinforcement and active learning;
• Recurrent and time-structure based methods;
• Neuro-inspired methods (including neuromorphic computing).

As an open access journal articles in MLST are published under a Creative Commons Attribution (CC-BY) licence and are freely available to everyone to read and reuse immediately upon publication. We believe that this is the most appropriate model in serving a research community where demand for open science principles is high. The journal is funded by article publication charges however IOP Publishing is covering all publication costs for any article submitted before the end of 2020. This means that our first authors will get all of the benefits that open access brings without paying a fee.

MLST also supports the principles of transparency and reproducibility in scientific research recognising that many research funders now require authors to make data related to their research available in an online repository. In this spirit authors in the journal will be required to include a data availability statement in their article and be encouraged to share their code and research data wherever it is appropriate and ethical to do so for the benefit of the research community.

Finally, a special mention for the members of the journal’s Editorial Board. MLST could not have launched without the significant contribution made by the members of the Editorial Board in writing, reviewing, and championing the journal amongst their research communities. Please visit the journal’s website to see a full list of Board Members: https://iopscience.iop.org/journal/2632-2153/page/editorial-board.

I hope that you enjoy reading our first content and that you will consider MLST for your own future research.

ORCID iDs

O Anatole von Lilienfeld  🐕 https://orcid.org/0000-0001-7419-0466

References

[1] Gordon O, Junquiera F and Moriarty P 2020 Embedding human heuristics in machine-learning-enabled probe microscopy Mach. Learn.: Sci. Technol. 1 015001
[2] Albarrán-Arriagada F, Retamal J C, Solano E and Lamata L 2020 Reinforcement learning for semi-autonomous approximate quantum eigensolver Mach. Learn. Sci. Tech. 1 015002
[3] Liu G, Chen M, Liu Y-X, Ayden D L and Cappellaro P 2020 Repetitive readout enhanced by machine learning Mach. Learn. Sci. Tech. 1 015003
[4] van der Oord C, Dusson G, Csányi G and Ortner C 2020 Regularised atomic body-ordered permutation-invariant polynomials for the construction of interatomic potentials Mach. Learn. Sci. Tech. 1 015004
[5] Cavaglia M, Gaudio S, Hansen T, Staats K, Szczepanczyk M and Zanolin M 2020 Improving the background of gravitational-wave searches for core collapse supernovae: a machine learning approach Mach. Learn. Sci. Tech. 1 015005
[6] Gopakumar V and Samaddar D 2020 Image mapping the temporal evolution of edge characteristics in tokamaks using neural networks Mach. Learn. Sci. Tech. 1 015006