Artificial Intelligence and Machine Learning for Nuclear Physics

Morten Hjorth-Jensen and many other

Facility for Rare Isotope Beams and Department of Physics and Astronomy, Michigan State University, MI 48824, USA
Department of Physics and Center for Computing in Science Education, University of Oslo, N-0316 Oslo, Norway

Advances in artificial intelligence/machine learning methods provide tools that have broad applicability in scientific research. These techniques are being applied across the diversity of nuclear physics research topics, leading to advances that will facilitate scientific discoveries and societal applications. This review provides a snapshot of nuclear physics research which has been transformed by artificial intelligence and machine learning techniques.

CONTENTS

I. INTRODUCTION

This Review represents an up-to-date summary of work in the application of artificial intelligence (AI) and machine learning (ML) in nuclear science, covering topics in nuclear theory, experimental methods, accelerator technology, and nuclear data.

Nuclear physics is a well-established field, with more than a century of fundamental discoveries covering a huge span of degrees of freedom, energy scales and length scales, ranging from our basic understanding of fundamental constituents of matter to the structure of stars and the synthesis of the elements in the Cosmos. Experiments produce data volumes that range in complexity and heterogeneity, thereby posing enormous challenges to their design, their execution, and the statistical data analysis.

Theoretical modeling of nuclear properties is, in most physical cases of interest, limited by the large amount of degrees of freedom in quantum-mechanical calculations. The analysis of experimental data and the theoretical modeling of nuclear systems aims, as is the case in all fields of physics, at uncovering the basic laws of motion in order to make predictions and estimations, as well as finding correlations and causations for strongly interacting matter. The broad aims of nuclear physics as a field correspond to a highly distributed scientific enterprise. Experimental efforts utilize many laboratories worldwide, each with unique operation, data acquisition, and analysis methods. Similarly, the scales of focus spanned in theoretical nuclear physics lead to broad needs for algorithmic methods and uncertainty quantification. These efforts, utilizing arrays of data types across size and energy scales, create a perfect environment for applications of AI/ML methods.

II. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING FOR NUCLEAR PHYSICS IN BROAD STROKES

Statistics, data science, and AI/ML form important fields of research in modern science. They describe how to learn and make predictions from data, and enable the extraction of key information about physical processes and the underlying scientific laws based on large datasets. As such, recent advances in AI capabilities are being applied to advance scientific discoveries in the physical sciences.

Ideally, AI represents the science of building models to perform a task without being explicitly programmed. ML tasks fall under the broader AI umbrella. We will henceforth refer to the methods discussed as "AI/ML". The idea is that there exist generic algorithms which can be used to find patterns in a broad class of datasets without having to write code specifically for each problem. The algorithm builds its own logic based on the data. The attentive reader should however always keep in mind that machines and algorithms are to a large extent developed by humans. The choice of a specific AI/ML algorithm is governed by the insights and knowledge about a specific system.

There exist many AI/ML approaches; they are often split into two main categories, supervised and unsupervised. In supervised learning, data are labeled and one lets a specific ML algorithm learn and deduce patterns in the datasets. This allows one to make predictions about future events and/or data not included in the training set. On the other hand, unsupervised learning is a method for finding patterns and relationship in datasets without any prior knowledge of the system. Many researchers also operate with a third category, dubbed reinforcement learning. This is a paradigm of learning inspired by behavioral psychology, where actions are learned to maximize reward. One may encounter reinforcement learning being accompanied by supervised deep learning methods. Furthermore, what is often referred to as semi-supervised learning, entails developing algorithms that aim at learning from a dataset that includes both labeled and unlabeled data.

Another way to categorize AI/ML tasks is to consider the desired output of a system. Some of the most common tasks are:

Classification: Outputs are divided into two or more classes. The goal is to produce a model that assigns inputs into one of these classes. An example is to identify digits based on pictures of hand-written numbers.

Regression: Finding a functional relationship between an input dataset and a reference dataset. The goal is to construct a function that maps input data into continuous output values.

Clustering: Data are divided into groups with certain common traits, without knowing the different groups beforehand. This AI/ML task falls under the category of unsupervised learning.

Generation: Building a model to generate data that are akin to a training dataset in both examples and distributions of examples. Most generative models are types of unsupervised learning.

In Table I we list many of the methods encountered in this work, with their respective abbreviations.

The methods we cover here have three central elements in common, irrespective of whether we deal with supervised, unsupervised, or semi-supervised learning. The first element is some dataset (which can be subdivided into training, validation, and test data), the second element is a model, which is normally a function of some parameters to be determined by the chosen optimization process. The model reflects our prior knowledge of the system (or lack thereof). As an example, if we know that our data show a behavior similar to what would be predicted by a polynomial, fitting the data to a polynomial of some degree would determine our model. The last element is a so-called cost (or loss, error, penalty, or risk) function which allows us to present an estimate on how good our model is in reproducing the data it is supposed to train. This is the function which is optimized in order to obtain the best prediction for the data under study. The simplest cost function in a regression analysis (fitting a continuous function to the data) is the so-called mean squared error function while for a binary classification problem, the so-called cross entropy is widely used, see, e.g., (Bishop, 2006; Hastie *et al.*, 2009; Murphy, 2012) for more details. We will henceforth refer to this element as the assessment of a given method.

Traditionally, the field of AI/ML has had its main focus on predictions and correlations. In AI/ML and prediction-based tasks, we are often interested in developing algorithms that are capable of learning patterns from existing data in an automated fashion, and then using these learned patterns to make predictions or assessments of new data. In some cases, our primary concern is the quality of the predictions or assessments, with perhaps less focus on the underlying patterns (and probability distributions) that were learned in order to make these predictions. However, in many nuclear physics studies, we are equally interested in being able to estimate errors and find causations. In this Colloquium, we emphasize the role of predictions and correlations as well as error estimation and causations in statistical learning and ML. For general references on these topics and discussions of frequentist and Bayesian methodologies, see, e.g., (Bishop, 2006; Goodfellow *et al.*, 2016; Hastie *et al.*, 2009; Murphy, 2012).

REFERENCES

C. M. Bishop (2006), *Pattern Recognition and Machine Learning* (Springer Verlag, Berlin).

I. Goodfellow, Y. Bengio, and A. Courville (2016), *Deep Learning* (The MIT Press, Cambridge, Massachusetts).

T. Hastie, R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (Springer Verlag, Berlin).

K. P. Murphy (2012), *Machine Learning: A Probabilistic Perspective* (The MIT Press, Cambdridge, Massachusetts).

TABLE I Table of AI/ML with indication on the main type of learning (S: supervised, U: unsupervised, Semi-S: semi-supervised).

Acronym	Method	Type of Learning
AE	Autoencoders	U
ANN	Artificial Neural Networks	S
BED	Bayesian Experimental Design	S
BM	Boltzmann Machines	U
BMA	Bayesian Model Averaging	S
BMM	Bayesian Model Mixing	Semi-S
BO	Bayesian Optimization	S
BNN	Bayesian Neural Networks	S
CNN	Convolutional Neural Networks	S
EMB	Ensemble Methods and Boosting, including Decision Trees and Random Forests	S
GAN	Generative Adversarial Networks	U
GP	Gaussian Processes	Semi-S
KNN	k-nearest neighbors	U
KR	Kernel Regression	S
LR	Logistic Regression	S
LSTM	Long short-term memory	S
PCA	Principal Component Analysis & Dimensionality Reduction	U
REG	Linear Regression	S
RL	Reinforcement Learning	Neither S nor U
RNN	Recurrent Neural Networks	S
SVM	Support Vector Machines	S
VAE	Variational Auto Encoders	U