Abstract

Machine learning (ML) techniques are widely used in science and industry as a powerful tool for data analysis and automation. Currently, in accelerator physics ML is represented as a young research field, demonstrating mixed results in the latest attempts. The presented work is devoted to exploration of appropriate ML methods for beam diagnostics. The target is to provide an overview of ML techniques which can be applied to improve beam diagnostics and general accelerator performance. Besides the results of ML tools currently used in modern accelerators and evaluation of these tools, we also demonstrate possible concepts with the potential for further investigation and give recommendations on efficient use of ML techniques in accelerators.

HISTORICAL MOTIVATION

Traditional Optimization Techniques

Various optimization problems arise in modeling and operation of accelerators. Multi-parameter optimization can be performed using well established methods such as simplex, as it was shown at KEKB [1] applied on minimization of vertical emittance in injector linac. Another examples for the application of simplex as optimization technique for accelerators can be found in [2], where the method was applied on tuning of beam delivery system in CLIC simulations.

Linear optics corrections using optimization algorithms to find a minimum beam size with multi-parameter knobs as input were performed already in 1993 [3]. Luminosity maximization and beam lifetime are typical multivariate optimization tasks also in circular colliders [4, 5].

For light sources, such methods as random-walk optimization are being used to reduce the vertical beam size [6]. Also online optimization using various measures of accelerator performance as objective functions is being successfully applied in operation [7]. An illustrating example towards applying ML is the development of machine based optimization using genetic algorithms, which have been used as well in accelerator design [8–10].

Limitations of Traditional Methods

The described examples demonstrate successful performance of traditional optimization tools in applications on linear optics corrections and problems with limited amount of optimization targets. Bigger challenges emerge when diagnostics of complex non-linear behavior is required and several variables have to be taken into account as final objective. The amount of time and computational power required by traditional methods might become unacceptable for future accelerators such HE-LHC and FCC. The main limitation of traditional optimization methods is that the objective function or specific rules and thresholds have to be known. In opposite, Machine learning (ML) methods can learn from given examples without requiring explicit rules.

RELEVANT MACHINE LEARNING CONCEPTS

ML techniques aim to build computer programs and algorithms that automatically improve with experience by learning from examples with respect to some class of task and performance measure, without being explicitly programmed [11].

Depending on problem and existence of learning examples, different approaches are preferred. If pairs of input and desired output are available, an algorithm can generalize the problem from the given examples and produce prediction for unknown input. ML algorithms that learn from input/output pairs are called supervised learning algorithms. Opposite to supervised learning, unsupervised learning algorithms solve the tasks where only input data is known. Unsupervised learning is suitable for the problems such anomaly detection, signal denoising, pattern recognition, dimensionality reduction and feature extraction. In the following a brief overview on significant machine learning concepts that can be used as supervised as well as unsupervised approaches is presented.

Artificial Neural Network

Artificial Neural Networks (ANNs) are well suited for learning tasks, where data is represented by noisy, complex sensor signals and the target output function may consist of several parameters. A basic ANN consists of a single processing unit (neuron), that takes the weighted inputs and an additional activation function to introduce the nonlinearity in the output. For more complex practical problems, ANNs are composed of several interconnected hidden layers with multiple neurons stacked.

ANNs can be used for both regression and classification problems. In case of classification the output can be either a class label or a probability of an item belonging to a class. The learning of ANN is performed using backpropagation algorithm [12] on a set of examples. For each example the training algorithm computes the derivatives of the output function of the network. The obtained gradients with respect to all weights are then used to adjust the weights in order to achieve a better fit to the target output. In backpropagation stochastic gradient descent or one of its improved extensions Adam [13] and AdaGrad [14] is applied as optimization method in order to minimize the loss between the network output values and the target values for these outputs by updating the connection weights.
One interesting property of the backpropagation based learning is its ability to discover useful intermediate representation inside the network. Features that are not explicitly given can be extracted using layers between input and output layers (hidden layers). Thus, properties of the input that are most important for the learning can be discovered. This ability is a great advantage of this method in contrast to ML techniques that use only predefined features. ANNs with many hidden layers called deep neural networks are able to use fewer neurons per layer and have a better generalization ability [15], however the optimization of these networks is not trivial. There are no strict rules for building ANN architecture (number of neurons, layers, initial weights) as it usually heavily depends on a particular problem. However, techniques to adjust the architecture parameters exist. A detailed overview on various ANN architectures and training methods and their suitability for different applications can be found in [16–18]. A broad presentation of concepts and applications of ANN to particle accelerators is given in [19].

Decision Trees and Ensemble Methods

Decision tree learning is a method for approximating discrete-valued target functions, which are represented by decision trees. Considering the case of classification, decision trees sort down the input instances from the root to leaf nodes. Usually, the splitting is based on one of the input parameters or a specified set of rules [20, 21]. Each leaf node corresponds to one class representing the most appropriate class label. For the regression the leaf nodes correspond to target values.

Using a single tree, a model might not be able to generalize and perform poorly on unexplored sample. One possible solution to overcome this problem is to build ensembles of trees [22]. By training several slightly different models and taking the average prediction, the variance of the model can be reduced.

Compared to ANNs, decision trees are simpler to interpret and to understand its way of obtaining the final results and the underlying process, e.g through the feature importance analysis. Feature importance analysis helps to understand the contribution of each input parameter to the correct decision during the training process. The ability of decision trees to evaluate the importance of input parameter is a significant advantage of these algorithms. Knowing the importance of the features we can reduce the model complexity and simplify the data preprocessing steps without significant accuracy loss.

Clustering

Cluster analysis includes methods of grouping or separating data objects into clusters, such that dissimilarity between the objects within each cluster is smaller than between the objects assigned to different clusters [23, 24]. Cluster analysis is used in a wide range of applications. Data clusters can be considered as a summarized representation of the data, such that group labels can describe patterns or similarities and differences in the data. Moreover, clustering can be used for prediction, such that classification of unseen data is performed based on knowledge about the properties of present data and by evaluating their similarity to the incoming data sample. The significant benefit of cluster analysis is the unsupervised learning approach, which means that no labeled data is needed to find a solution.

The simplest and the most commonly used clustering algorithm is k-means [25], which is based on centroid search. Another kind of clustering algorithms are the density-based algorithms such DBSCAN [26], that views clusters as areas of high density separated by areas of low density, instead of looking for the centroids. Decision tree based methods also can be applied for cluster analysis using the data splits based on different features. Most of cluster analysis techniques allow to build clusters in a multidimensional space.

Apart from classification and pattern recognition, cluster analysis can be used as denoising method looking for abnormalities in the signal. Moreover, building clusters combining a large set of different observables can simplify the data visualization and manual analysis, such elimination of outliers in the measurements and detection of anomalies.

OVERVIEW ON CURRENT APPLICATIONS

Meeting the demand of experimenters induce various challenges for accelerator design in general and in particular for beam control and diagnostics. Considering impressive results of ANNs applied in various scientific fields [27–31], among others in HEP [32] and the increase in available computational power, ML can cast the light on novel solutions for these challenges. In the following we demonstrate some ML applications currently being used in accelerator technology and ongoing research on potential ML based approaches.

Optimization and Prediction

A complex system such as an accelerator, which beam dynamics exhibits nonlinear response to machine settings can be considered as a typical ML task. ML methods are especially suited for non-linear and time-varying systems with large parameter spaces. Due to the constant increase of machine design complexity and development of new interacting systems, traditional techniques might become insufficient.

ANN based application has been successfully applied at the Linac Coherent Light Source (LCLS) to predict x-ray pulse properties using electron beam and x-ray parameters as input [33]. The method is general and can be applied at any XFEL facility to obtain fast x-ray diagnostics. ANN is able to decode complex hidden correlations between parameters obtained from slow diagnostics such as photon energy and properties measured by fast diagnostics. For the training, small samples obtained at low repetition rate are used to predict complex diagnostics at a high repetition rate.

Similar approach has been applied at TEU-FEL by training an ANN with controllable machine settings such RF power and solenoid strengths as inputs in order to produce
prediction of electron beam parameters such beta functions and emittance [34].

A special kind of ANN, convolutional neural networks (CNN) [35] have been applied at FAST on image based diagnostic during beam operation [36]. A combination of a CNN and a feed-forward NN yields promising results for the prediction of beam parameters on simulated datasets. The model uses simulated cathode images, solenoid strengths and the gun phase as inputs and produces a prediction for various downstream beam parameters.

Extremum Seeking (ES) technique [37, 38] in combination with ML is being applied to accelerator tuning and optimization providing promising results. In this application ANN is used for fast tuning in order to obtain a close approximation for the target settings. ES is used for beam-based adaptive feedback to track the actual time-varying optimal parameter settings. The advantages of this method are the model-independent approach, the ability to perform the tuning on many coupled parameters and handle time-varying noisy data.

Another example concerning the prediction of beam parameters is the application of various ML algorithms to the problem of inferring the actual beam profile width from measured profiles that are distorted by space-charge effects [39]. The promising results obtained on simulations data show the potential of the method to be further investigated on real measurements in order to reduce cost of Ionization Profile Monitors.

**Automation of Operation**

Often there is a need to introduce various experiment-driven settings. In this case a large amount of free parameters has to be taken into account in order to meet the requirements and find optimal machine settings under given limitations. Only few parameters can be processed by a human at once and it is not feasible to produce forecasts taking into account all possible factors and correlations. Moreover, humans can perform differently on the tasks where the decisions can be subjective, which might lead to significantly different results. Also the automation of routine tasks could bring significant improvements into daily operation, such that the focus of operation can be transferred to complex tasks and rare events that require an expertise.

Apart from ANN, it is also possible to apply other kinds of regressors or classifiers in accelerator control such Gradient Boosting classifier [40] as it was shown in beam loss pattern classification for LHC [41]. The beam loss maps performed in controlled conditions are used in order to train a model to classify the type of losses during the LHC machine cycle.

Another example of ML based automation is the automatic alignment of collimators in SPS and LHC [42]. The method computes optimum angular settings for the collimators without human intervention. Fully automated alignment was achieved after the introduction of the ML based detection of alignment spikes in the losses which are used in order to determine if a collimator needs to be moved.

**Lattice Imperfections Correction**

Attempts to build beam diagnostics and beam control systems using ML have been made already in the past decades [43–45]. The application described in [45] was built to detect dipole errors aiming to develop rapid commissioning. In this application, the dipole errors are obtained from the deviation of the measured beam position from the computed one. The simulations and tests have been performed on a relatively small machine (8 FODO cells, 8 BPMs), however the accuracy of the trained models decreased significantly after introduction of more than 2 dipole field defects. Anyway, given the early stage of ANN technology in that time, the obtained results have shown the potential of ML solution to be applied in beam control systems.

Recently, we applied ML on optics correction at LHC to predict the correction knobs settings required to cancel the quadrupole field errors [46,58]. In this case, optics correction is defined as a regression problem that can be solved by training a model using past measurements or simulations and its corresponding corrections (supervised learning).

In order to create a training set, random errors are introduced into the MADX-variables that represent physical circuits. In real measurements the optics errors are caused...
Transfer Learning

This method is fully integrated into optics measurements at LHC and has been successfully used during commissioning re-trained on a very small data set [51,52]. Thus, small amount of real optics measurements and performed corrections can be used to tune the model trained on simulations to significantly increase the model quality.

Anomaly Detection

Anomaly detection techniques are suitable for the detection of unusual events that do not conform to expected patterns. They also can be used to identify outliers and remove noise. Anomaly detection can be performed using classification on labeled data (supervised learning), cluster analysis (unsupervised learning) or applying semi-supervised learning methods such as autoencoder, a special ANN representing the model trained on normal data set and then detect the anomalies based on the value of the loss function generated by the representative model on the given test sample [53].

An early example on anomaly detection in beam diagnostics in storage ring (Pohang Light Source) is the application of ANN to predict the orbit at particular BPM based on measurements at other BPMs [54]. A large deviation between measured and predicted orbit should mark malfunctioning BPM.

An example for anomaly detection using cluster analysis is the detection of faulty BPMs at LHC based on harmonic analysis on turn-by-turn data [55]. The main issue regarding the problem of faulty BPMs is the appearance of unphysical data in the reconstructed optics functions. Most of the noise can be removed using traditional methods based on SVD and FFT [56] and manual cleaning, however, faulty data samples still can be observed in the optics functions. Since no labeled data is available and the data analysis has to be performed on multidimensional parameter space, clustering appears as an effective solution. The analysis as shown in Fig. 3 is performed on a three dimensional parameter space containing the betatron tune, the amplitude of the measured oscillations and the noise to amplitude ratio. The data is normalized to the range [0,1] and separated into IR and Arcs BPMs due to the different data points distribution in these regions.

Since the appearance of outliers in the data affects the computation of the mean of parameters, the algorithms based on centroids search such as K-means are not appropriate for our problem. Instead of centroid-based algorithms, DBSCAN and Isolation Forest (IF) [57] have been applied on the turn-by-turn data after harmonic analysis. Compared to the application of DBSCAN [58] better results could be achieved using IF algorithm. IF uses an ensemble of randomized decision trees averaging the path lengths, which represent the number of splits required to isolate a data point. Shorter paths are produced for anomalies. Due to the randomization and combination of several decision trees, the method performs better than single-model methods. Another advantage of IF is that the algorithm requires only the number of trees and the proportion of outliers in the data set (threshold for the decision function) as input parameter.

This method is fully integrated into optics measurements at LHC and has been successfully used during commissioning...
ML is well known for surpassing human performance in some specific tasks such fraud detection, forecasting of market trends and risks, online recommendations, recognition of voice and images and in general in discovering correlations in large scale datasets. Most of the named tasks can find analogies in beam control and diagnostics. For example, anomaly detection methods applied for fraud detection can be used to detect defects in the instrumentation and forecasting techniques can be transferred to predict beam behavior during operation.

Typical characteristic of supervised ML tasks is the ability to deal with large amount of structured data. This leads to the conclusion that the implementation of supervised ML solutions requires large existing training datasets or development of appropriate data acquisition tools in order to provide the data in "machine-understandable” format, which is not necessarily available out-of-the-box since the traditional control systems usually imply human intervention. The effort that has to be put on automation such as building data acquisition infrastructure and training of complex models might be more costly and resources expensive than traditional methods. On the other hand, automation of some particular systems using ML as it was done for collimators alignment at LHC [42] is very effective and can save operational resources. Diagnostics of beam losses and beam optics corrections can also benefit from ML as it was shown in [41] and [58], since these methods rely on already existing data from dedicated measurements.

The ability of unsupervised learning to discover unknown patterns in the data is useful especially for anomaly detection tasks such as detection of instrumentation defects, e.g. using clustering for faulty BPMs signal. Such methods can be performed directly without training in arbitrary accelerator systems.

**Further Applications**

Besides presented applications, another field where we can potentially benefit from ML are highly complex simulation tasks which demand large computational power or even are not solvable using the available resources. The predictive power of ANNs could offer an alternative solution for problems such as dynamic aperture computation avoiding costly turn-by-turn simulations. Moreover, different ML methods can be potentially used for maintenance of accelerators, such that system defects can be predicted based on deviations from expected performance and required intervention can be performed prior to actual failures [59]. Beam lifetime optimization is another task where ML techniques can be potentially applied [60]. Such methods as decision trees and autoencoders appear suitable for analysis of beam parameters correlations and their importance for beam lifetime prediction and maximization.
REFERENCES


9. Machine parameters measurements and others
9. Machine parameters measurements and others