ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB184

UNSUPERVISED LEARNING TECHNIQUES FOR TUNE CLEANING MEASUREMENT

H. Garcia-Morales*, University of Oxford, United Kingdom
E. Fol, R. Tomás García, CERN, Geneva, Switzerland

Abstract

Precise measurements of tune and its stability are crucial for various optics analyses in the LHC, e.g. for the determination of the β^* using K-modulation. LHC BBQ system provides tune measurements online and stores the tune data. We apply unsupervised machine learning techniques on BBQ tune data in order to provide an automatic outlier detection method for better measurements of tune shifts and unexpected tune jitters.

INTRODUCTION

Anomaly detection is one of the main applications of Machine Learning techniques in many different areas. Correct identification of rare items, events or observations which are suspicions by differing significantly from the majority of the data is essential to obtain more reliable and realistic results of our observables.

In the accelerator environment, the tune Q is an essential quantity related to many other quantities relevant to evaluate the performance of the machine. The LHC BBQ system provides continuous tune measurement. However, the BBQ system is not able to automatically identify measurement outliers due to faulty BPMs or wrong data acquisition which are quite common during machine operation. These outliers introduce a significant tune uncertainty which translates into imprecise measurements of derived quantities.

In general terms, for the LHC and most importantly for its future upgrade, the HL-LHC, it is desirable to keep the tune uncertainty below 10^{-5} in order to minimise the luminosity imbalance between high-luminosity experiments, ATLAS and CMS [1, 2]. Unfortunately both simulations and measurements are presently exceeding this limit, therefore it is important to find new strategies to reduce tune uncertainty to tolerable limits.

In this paper we show the qualitative progress towards a better tune cleaning measurement using machine learning techniques. First, the methodology used is introduced. Secondly, the different algorithms tested are briefly explained. Finally, the different algorithms have have been been tested in different regimes, and their performance for tune measurement cleaning, are evaluated.

ANALYSIS APPROACH

Usually, the tune signal is noisy and some outliers are clearly visible. In addition, when changing between different optics configurations, the working point is in most cases altered. However, sudden changes in the working point are not always easily predicted and might be due to some disturbance such as unknown machine configuration changes [3]. Therefore, cleaning the tune time series is not always straightforward, in particular when there is a change in the working point. A possible strategy would be to treat the data by cleaning the different segments associated to the different working points, but this means that we need to re-adapt our cleaning algorithm every time. Therefore, it makes the cleaning process difficult to generalize using classical tools.

Here we consider a different approach. Instead of cleaning the time series, the idea is to act on the tune space (Q_x, Q_y) applying clustering algorithms in order to distinguish noise from signal and at the same time to identify the different working points present during the data acquisition process. In the next section the different clustering algorithms considered are briefly introduced.

CLUSTERING ALGORITHMS

Clustering is one of the most popular machine learning techniques for data classification and anomaly detection. Among others, k-means is broadly used for data clustering. However, it requires to know the number of clusters beforehand which is impossible or very difficult to predict in the case of tune measurement. For that reason, k-means is not considered for tune cleaning and other algorithms are being explored.

Density Based Scan

Density based scan (DBSCAN) is a clustering algorithm [4] which, given a set of points in some space, groups together points that are closely packed together, marking as outliers points that lie alone in low-density regions. The two input parameters we need to include are the maximum distance between two samples ϵ for one to be considered as in the neighborhood of the other and the number of samples N in a neighborhood for a point to be considered as a core point. The particular characteristics of this algorithm makes it very suitable for tune cleaning.

Local Outlier Factor

The local outlier factor (LOF) algorithm is based on a concept of a local density [5], where locality is given by k nearest neighbors, whose distance is used to estimate the density. By comparing the local density of an object to the local densities of its neighbors, one can identify regions of similar density, and points that have a substantially lower density than their neighbors (outliers).

* hector.garcia.morales@cern.ch

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB184

Isolation Forest

Isolation Forest (IF) is an unsupervised algorithm for anomaly detection which works on the principle of isolating anomalies [6]. The contamination parameter takes into account the expected number of outliers in the sample. In the case of tune measurements, we expect approximately 10% of the tune measurements to be outliers.

ALGORITHM COMPARISON

The performance of the three ML algorithms presented above are qualitatively compared in different environments where the tune had different behaviours.

Quiet Period

As a first example we consider the case of a quiet period where the tune does not suffer any sudden change but only contains some noise. As expected, in this case, all three methods perform similarly. However, due to the simplicity of the set up of its input parameters, DBSCAN is preferred. The results of the clustering using DBSCAN are shown in Fig. 1 where data and noise are correctly classified.

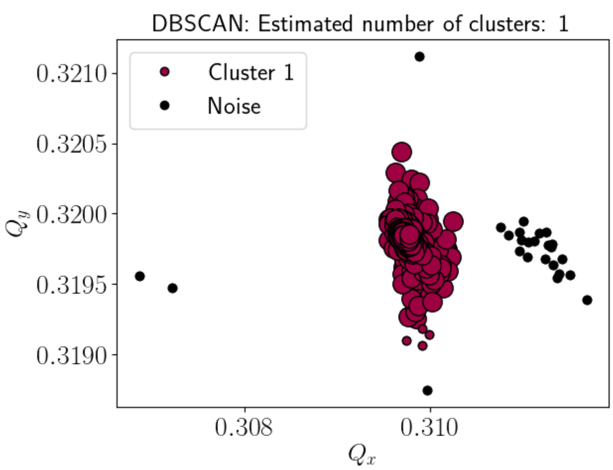

Figure 1: Clustering performed in a regular tune measurement in the LHC without major disturbances. DBSCAN automatically determines the number of clusters.

Tune Jumps

Sometimes, the tune suffers some expected or unexpected jumps [3]. In this situation, classical cleaning tools are not sufficient to clean, for instance, points measured during the transition from one working point to the next as shown in Fig. 2. In this case we observed that the best algorithm is again DBSCAN. Not only because one can intuitively set the model parameters, but because is the only one that classifies the different clusters and allows an easy analysis of each of these clusters separately. The uncertainty in the tune measurement in each of these clusters is reduced to around 5×10^{-5} without fully optimising the algorithm. IF and LOF are also correctly classifying signal from noise but they do not automatically classify the different clusters.

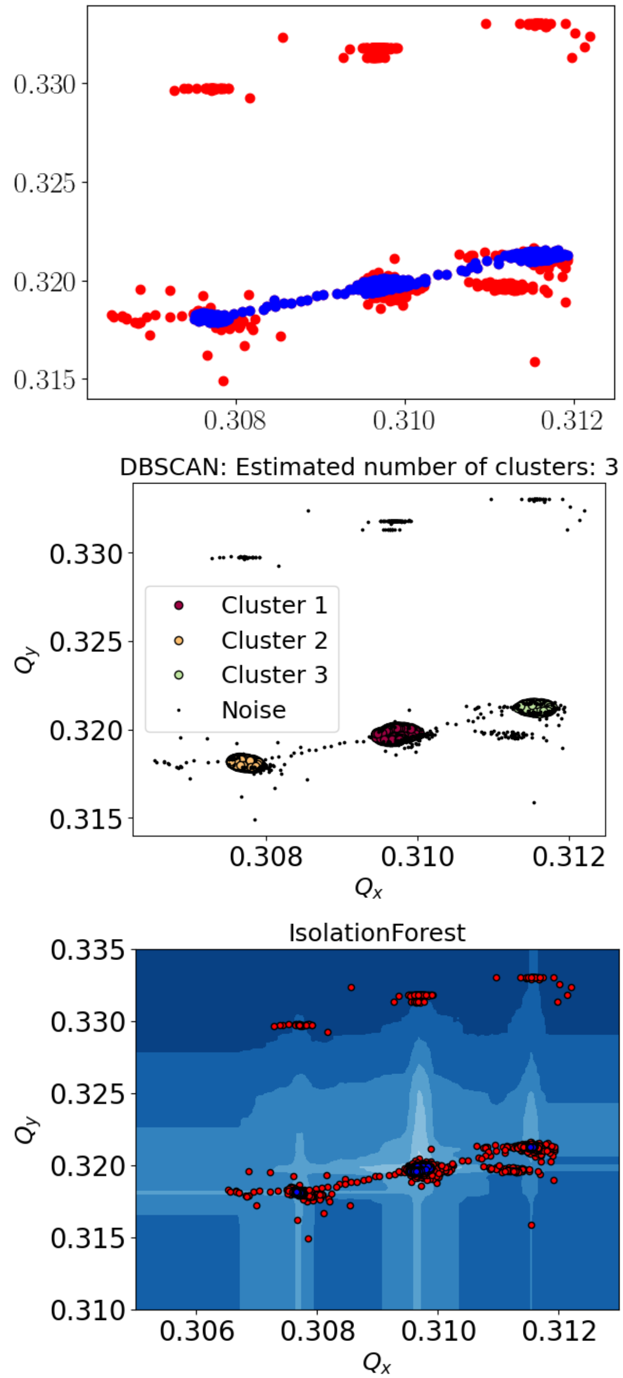

Figure 2: Comparison of different techniques for clustering and noise removal. Classical cleaning (top); DBSCAN ($\epsilon = 0.7 \times 10^{-4}$ and $N_n = 70$) (middle); and isolation forest (0.2 contamination) (bottom).

K-Modulation

The currently preferred method to determine the β^* in colliders such LHC, HL-LHC, FCC-hh and SuperKEKB is K-modulation [1, 7–10]. This method relies on the modulation of the gradient of the quadrupoles closest to the IP. The induced tune shifts allow to determine the average β

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB184

function in the quadrupoles. The waist shift w and β^* can then be calculated via propagation.

In this case, since the relationship between Q_x and Q_y is approximately linear, classical tools could be useful to correctly classify signal from noise. However, this linear relationship is not always ensured and some deviations from the linear regime may lead to an incorrect classification of signal and noise. For that reason, DBSCAN is also preferred for performing a correct classification, as shown in Fig. 3.

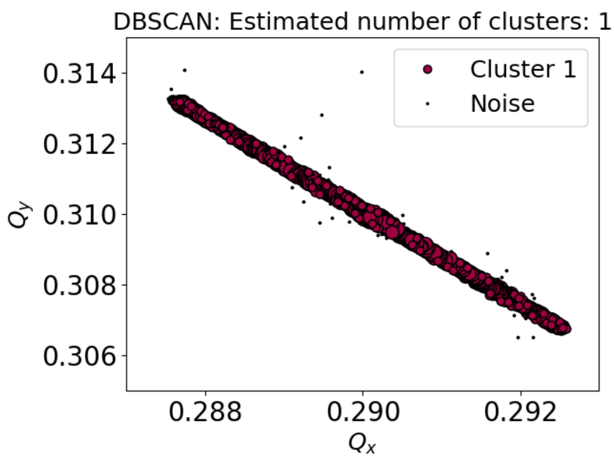

Figure 3: Tune cleaning during K-modulation using DBSCAN algorithm ($\epsilon = 10^{-4}$ and $N_n = 50$).

Closest Tune Approach

Due to coupling between the two transverse planes, the horizontal and vertical tunes have a minimum separation which is determined by the amount of coupling present in the machine. Coupling is sometimes measured by sequentially adjusting both transverse tunes to the same fractional value. Due to the presence of coupling, there is a minimum tune separation.

During coupling measurement, knobs are used to sequentially reduce the tune separation. At each step, the tune slightly changes. With classical cleaning tools, the tune must be cleaned at each step. However, with clustering algorithms, in principle, one should be able to correctly identify each step and to globally distinguish signal from noise. For this task, DBSCAN seems to be a promising tool. However, tuning the algorithm parameters to perform a global cleaning is not straightforward and more studies are required to find optimal values. In Fig. 4, the result of applying DBSCAN during closest tune approach, and three step period is shown. We can see how the algorithm is able to correctly identify the clusters that correspond to the different steps.

FUTURE STUDIES

The performance of clustering algorithms for correct identification of outliers in tune measurements can be expanded including more features that can be relevant for a better classification. Future studies may include some sort of automatisation for the parameter selection as well a more quantitative

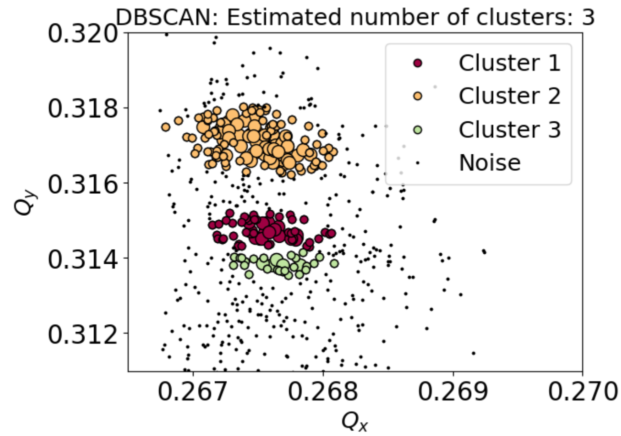

Figure 4: Tune measurement cleaning using DBSCAN during closest tune approach measurement showing three clusters corresponding to three different steps ($\epsilon = 3 \times 10^{-4}$ and $N_n = 22$).

analysis of the performance of these algorithms as well as a proper evaluation of the tune uncertainty once optimal parameters are found and cleaning is performed.

CONCLUSIONS

This is the first step towards a better measurement of tune in order to reduce its uncertainty using Machine Learning techniques. Different algorithms have been compared to traditional cleaning techniques. It was shown that in situations where the tune experiences sudden changes, or when the relationship between the Q_x and Q_y is non-linear, machine learning tools are a more flexible and reliable tool for cleaning the tune measurement with respect to classical cleaning tools. Among the different ML tools, we have compared density based algorithms and anomaly detection using algorithms such as Isolation Forest, with DBSCAN being the one that performs best, while at the same time as as being the most intuitive. Isolation Forest also performs very well in some scenarios, although its implementation is less intuitive. This new approach can be easily implemented in other accelerators.

REFERENCES

[1] T. Persson *et al.*, “LHC optics commissioning: A journey towards 1% optics control”, *Phys. Rev. ST Accel. Beams*, vol. 20, p. 061002, 2017. doi:10.1103/PhysRevAccelBeams.20.061002

[2] F. Carlier *et al.*, “Optics Measurements and Correction Challenges for the HL-LHC”, CERN, Geneva, Switzerland, Rep. CERN-ACC-2017-0088, 2017.

[3] S. Visavadia, “Investigation into Possible Sources of Tune Jitter in the LHC”, Ph.D. thesis, CERN, Geneva, Switzerland, 2019.

[4] M. Hahsler, M. Piekenbrock, and D. Doran, “dbscan: Fast Density-Based Clustering with R”, *Journal of Statistical Software*, vol. 91, no. 1, pp. 1–30, 2019. doi:10.18637/jss.v091.i01

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-MOPAB184

[5] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF", *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000. doi:10.1145/335191.335388

[6] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation Forest", in *Proc. 2008 Eighth IEEE International Conference on Data Mining (ICDM'08)*, Pisa, Italy, Dec. 2008, pp. 413-422. doi:10.1109/ICDM.2008.17

[7] E. H. Maclean *et al.*, "New approach to LHC optics commissioning for the nonlinear era", *Phys. Rev. Accel. Beams*, vol. 22, p. 061004, 2019. doi:10.1103/physrevaccelbeams.22.061004

[8] F. Carlier and R. Tomás, "Accuracy and feasibility of the β^* measurement for LHC and High Luminosity LHC using k-modulation", *Physical Review Accelerators and Beams*, vol. 20, no. 1, p. 011005, Jan. 2017. doi:10.1103/physrevaccelbeams.20.011005

[9] M. Hofer *et al.*, "K-Modulation in Future High Energy Colliders", in *Proc. 10th Int. Particle Accelerator Conf. (IPAC'19)*, Melbourne, Australia, May 2019, pp. 476-479. doi:10.18429/JACoW-IPAC2019-MOPMP022

[10] P. Thrane *et al.*, "Measuring β^* in SuperKEKB with K Modulation", *Phys. Rev. Accel. Beams*, vol. 23, p. 012803, Jan. 2020. doi:10.1103/physrevaccelbeams.23.012803