ISBN: 978-3-95450-208-0 doi:10.18429/JACoW-IPAC2019-WEPGW081

UNSUPERVISED MACHINE LEARNING FOR DETECTION OF FAULTY BEAM POSITION MONITORS

E. Fol*, J. Coello de Portugal, R. Tomás, CERN, 1211 Geneva 23, Switzerland
*also at Johann-Wolfgang Goethe University, 60438 Frankfurt am Main, Germany

Abstract

Unsupervised learning includes anomaly detection techniques that are suitable for the detection of unusual events such as instrumentation faults in particle accelerators. In this work we present the application of a decision trees-based algorithm to faulty BPMs detection at the LHC. This method is fully integrated into optics measurements at LHC and has been successfully used during commissioning and machine developments (MD) for different optics settings in 2018.

INTRODUCTION

In the LHC there are 524 beam position monitors (BPMs) per plane and per beam. Optics measurements are mainly concerned by phase advance measured between BPMs [1]. The phase advances are inferred from a harmonic analysis of the turn-by-turn transverse beam motion recorded by the BPMs around the ring. Faulty BPMs produce unreliable signal and hence reduce the quality of the optics measurements. The identification of anomalies in acquired beam position measurements requires the application of automatic tools as well as human intervention. Most of the noise and faulty signals can be removed using predefined thresholds for recorded signal, as well as through applying advanced signal-improvement techniques based on SVD [2] to reduce noise in BPM readings. However few nonphysical values are still observed in reconstructed optics, this combination of numerical techniques and basic threshold cuts appear not to be sufficient. Therefore, alternative techniques are required. Not all reasons for the appearance of BPM anomalies are known, therefore we cannot define new thresholds which would indicate faulty BPMs. The required cleaning method should not rely on numerical cuts or "learn" from the results of existing tools. Looking for a machine learning based solution for the described problem, we are going into the domain of *unsupervised learning*.

UNSUPERVISED LEARNING

Unsupervised learning deals with tasks where only input data is available and the target is to find patterns in the given data or to extract new information. Opposite to supervised learning, unsupervised techniques provide the possibility to identify unusual patterns (outliers) without being trained on labeled data. Several unsupervised learning algorithms have been applied to optics measurements at LHC [3,4], among others the Isolation Forest (IF) algorithm [5]. It detects anomalies using binary trees to isolate each data point. The *anomaly score* assigned to each data point corresponds to the number of splits taken to isolate this point, averaged over the number of trees.

* elena.fol@cern.ch

EXPERIENCE WITH ISOLATION FOREST APPLICATION AT THE LHC

Due to experience with the systematical observation of few non-physical “spikes“ in the reconstructed optics functions, we assume that only a small fraction of bad BPMs is remaining after the cleaning (SVD and numerical cuts). IF requires the proportion of outliers (contamination) in the data as input of the algorithm. Initially, the contamination was set to 1% in the arcs and 2.5% in the IRs. The tuning of this contamination parameter on simulated BPM faults will be discussed in the next section.

The parameters that are considered significant for bad BPMs identification are the betatron tune, the amplitude of the measured oscillations and the noise scaled with the amplitude of the signal. During commissioning and MDs in 2018 the new cleaning method was used complementary to the existing techniques. In case non-physical outliers were observed in optics functions, the harmonic analysis data was additionally cleaned with IF and the optics analysis was repeated. Figure 1 shows a comparison between the beta-beating reconstructed from the measurements before and after applying IF. It can be clearly observed that most of the remaining outliers have been removed.

Figure 2 shows the summary on IF results on several measurements in 2018 performed for both beams for different optics settings. It can be observed that most of the spikes remaining after SVD and thresholds-based cleaning could be eliminated by IF. It has to be noted that there is not necessar-

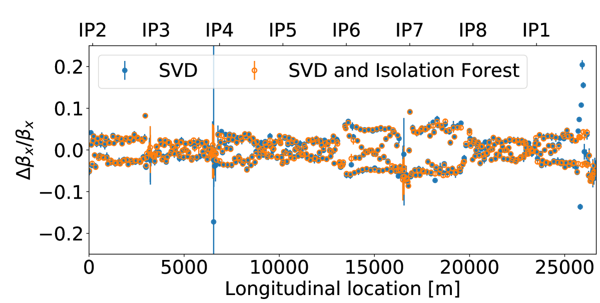

Figure 1: The comparison between beta-beating computed before and after IF cleaning demonstrates that IF anomaly detection significantly reduces the number of unphysical spikes and reduces the size of errorbars. The optics is computed for Beam 2 in horizontal plane with β^*=50 cm.

ISBN: 978-3-95450-208-0 doi:10.18429/JACoW-IPAC2019-WEPGW081

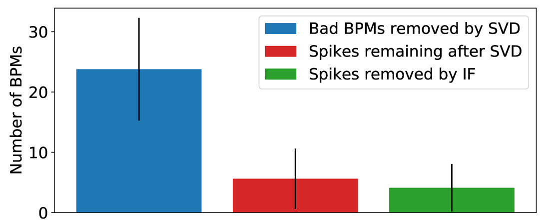

Figure 2: This summary represents the number of outliers in β-beating and phase advance averaged over 10 measurements of Beam 1 and Beam 2 performed during MDs in 2018.

ily a direct relation between the location of the spikes and the actual bad BPMs. The β-functions are calculated from the phase advances between BPMs in a certain range [6, 7]. Due to the use of a range of N BPMs, a single faulty BPM may cause multiple spikes in the optics and the produced spikes might appear not directly at the position of the bad BPM. Since the definition of a spike is subjective, it is not possible to conclude on the exact amount of actually faulty BPMs that are removed and the number of good BPMs that are wrongly recognized as faulty. Since the knowledge about actual defective BPMs is not available, the assessment of cleaning algorithms has to be performed on simulations where the actual bad BPMs are known and can be used as labeled data to evaluate the performance of the method.

PERFORMANCE EVALUATION ON SUPERVISED DATASET

First, turn-by-turn signal is generated for 6600 turns without any perturbation using ion optics with β^*= 50 cm in IP1, 2 and 5. Every BPM is given 0.1 mm Gaussian background noise. In the second step, the signal of randomly chosen BPMs is artificially perturbed - these BPMs have to be identified as bad. In real measurements, the reasons for the appearance of faulty signal are unknown, but there are specific artifacts which are known to be related to faulty BPMs. Considering the BPMs removed by traditional tools and the remaining spikes, we can conclude that around 5.5% of measured positions are erroneous. Hence, we perturb 5.5% of original simulated turn-by-turn data. The following perturbations are used to introduce BPM faults:

- Gaussian noise with σ=0.3 mm is added to the signal
- Signal is replaced by a random value in range [-20, 20] in a single turn
- Tune of the signal deviates by 10^{-5} from the rest
- Flat zero signal
- Signal is replaced by zero in a single turn
- Multiple failures (tune deviation, random value in one turn and noise) are present

2 BPMs with flat zero signal and 5 BPMs of each remaining failure type are introduced in each simulated measurement producing 27 bad BPMs per plane in total. As in real optics measurements, we first clean the simulated BPM signal with existing tools using default settings for SVD [2] and signal thresholds as shown in Table 1.

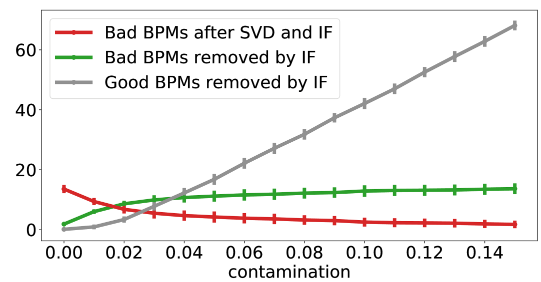

Figure 3: Adjustment of contamination factor of IF algorithm and its effect on the trade-off between cleaning of bad BPMs and removal of good ones.

Table 1: Cleaning Thresholds Used in 2018. The SVD settings are described in [2].

SVD		**Signal cuts**	
SVD cut	0.925	min peak-to-peak	10^{-5} mm
SVD modes	12	max peak-to-peak	20 mm
		tune deviation	10^{-5}

Knowing from the simulations that around 12 bad BPMs are removed by traditional cleaning tools and 15 bad BPMs are remaining, the contamination should be set to $15/(524-12) \approx 0.029$. To study the influence of the contamination parameter on the results, we run IF multiple times increasing the contamination number from 0 to 0.15 step-wise. Figure 3 illustrates the trade-off between eliminating bad BPMs and removing good BPMs as side effect. Based on the results we conclude that the optimal contamination factor lies around 0.02 as expected.

The results summarized in Fig. 4 show that the amount of actual good removed BPMs is small compared to the amount of identified faults. It must be noted that the presence of faulty BPMs has a larger negative effect than few missing BPMs due to the optics computation method [7]. Compared to the experience gained applying the IF algorithm on measurements in 2018, in simulation we observe more bad BPMs that are remaining after applying traditional tools than spikes remaining in real measurements. To be noted that in measurement summary shown in Figure 2 we consider subjective observation of nonphysical spikes in beta-beating and total phase as bad BPMs. Possible explanation for the difference in the number of good removed BPMs between measurements and simulations is that not all of bad BPMs introduced into simulations might cause a spike in the optics. Also, not all fault artifacts are known and included into the simulations. Another possible reason is that the contamination factor used in IF algorithm in 2018 differs from the actual observations in 2018. As shown in Fig. 2 the fraction of detected bad BPMs decreased to 5.5% compared to the statistics from the past [3] where 10% of BPMs in each plane were identified as bad.

ISBN: 978-3-95450-208-0 doi:10.18429/JACoW-IPAC2019-WEPGW081

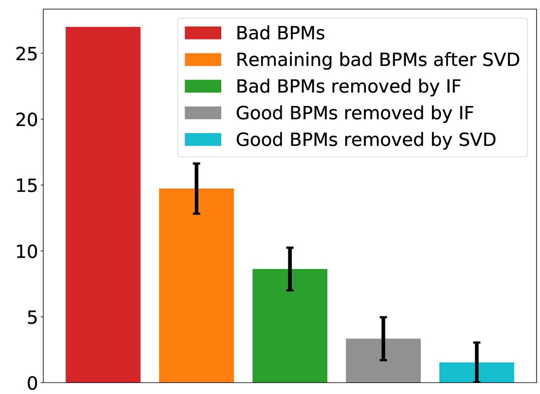

Figure 4: Averaged results of faulty BPMs detection on 20 simulated measurements using contamination factor 0.02.

COMPARISON TO CLUSTERING

Different clustering algorithms have been considered to detect faulty BPMs at the LHC, complementary to the existing tools. A possible solution is to apply density-based algorithms such as DBSCAN [8] which views clusters as areas of high density separated by areas of low density. DBSCAN has been tested offline on real LHC turn-by-turn data. The results of DBSCAN demonstrated improvements on data cleaning [9], however a significant amount of outliers remained in measured optics functions.

Local Outlier Factor (LOF) algorithm measures the local deviation of density of a given sample with respect to its neighbors [10]. Like DBSCAN, LOF improves the quality of turn-by-turn data resulting in less outliers in computed optics functions. From the experience of applying different algorithms on turn-by-turn data from past measurements, we observe that the optics reconstructed from the data cleaned with IF shows less unphysical spikes compared to the other two methods [11]. In order to examine the performance and suitability of each method for faulty BPMs detection, we carry out the simulation procedure as described in previous section. DBSCAN, LOF and IF are compared as complementary methods to the traditional cleaning tool for optics measurements at the LHC. Figure 5 demonstrates the result of elimination of faulty BPMs remaining after the application of traditional cleaning tools. The comparison shows nearly identical performance of LOF and IF algorithms on simulated BPM signal. The anomaly score in LOF method depends on how isolated the object is with respect to the surrounding neighborhood, which is very similar to IF algorithm. The locality is given by k-nearest neighbors, whose distance is used to estimate the density in the neighborhood and therefore, the number of nearest neighbors has to be specified as input parameter of the algorithm. Since the structure of the measurements data can vary significantly depending on the BPM location and machine settings, a general valid definition of local neighborhood becomes problematic. Due to the randomization and combination of several decision trees, IF algorithm should be more robust to deviations in the data structures than single-model methods [12] and

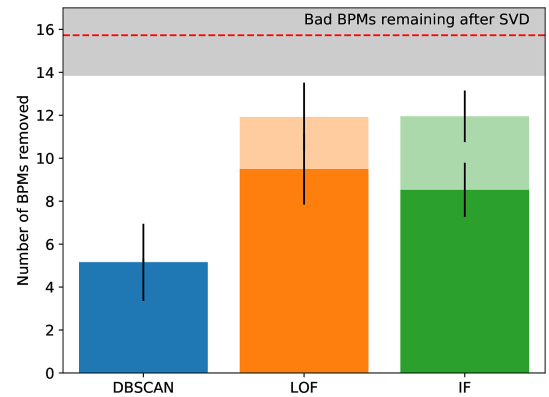

Figure 5: The comparison is carried out on 20 simulations for each plane, the results are averaged. Each bar represents the number of BPMs removed by the method. Dark fraction corresponds to the number of removed BPMs that are actually bad.

hence it was preferred to other clustering techniques. Another advantage of IF is that it requires only the number of trees and contamination of the data set as input parameters allowing simpler tuning of the algorithm and more general application.

In case of large machines such as the LHC equipped with hundreds of BPMs, it is important to decrease the number of faulty signal artifacts as much as possible, because a single faulty BPM can affects the optics computation at multiple locations. The absence of few good BPMs that might be caused by IF algorithm does not have a significant negative effect since the optics computation can be propagated to the next available BPM. Considering smaller machines, it is crucial to keep as much BPM information as possible, removing only critically erroneous signal. In this case, a method such as DBSCAN appears to be more appropriate since, as it was shown on simulations, the method does not identify any good BPMs as faulty, however a portion of bad BPMs is still remaining in the measurement.

CONCLUSION

The presented study demonstrates the ability of IF algorithm to successfully identify anomalies in measurement data caused by BPM faults without significant loss of good signal. Due to the randomization and combination of several decision trees, the method performs better than clustering techniques and does not depend on signal-specific thresholds compared to other techniques already used in optics measurements. Summarizing the results of application of IF algorithm in 2018, we can observe the reduction of nonphysical spikes in the optics achieved by IF. The new method complements the existing techniques in an efficient way, such that most of remaining non-physical outliers are eliminated, without affecting the optics computation negatively by removing a small fraction of good BPMs.

ISBN: 978-3-95450-208-0 doi:10.18429/JACoW-IPAC2019-WEPGW081

REFERENCES

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

[2] R. Calaga, R. Tomás, "Statistical analysis of RHIC beam position monitors performance", *Phys. Rev. ST Accel. Beams* vol. 7, p. 042801, 2004, doi:10.1103/PhysRevSTAB.7.042801

[3] E. Fol, "Evaluation of Machine Learning Methods for LHC Optics Measurements and Corrections Software", Master thesis, CERN-THESIS-2017-336.

[4] E. Fol, F. S. Carlier, J. M. Coello de Portugal, A. Garcia-Tabares, and R. Tomás, "Machine Learning Methods for Optics Measurements and Corrections at LHC", in *Proc. 9th Int. Particle Accelerator Conf. (IPAC'18)*, Vancouver, Canada, Apr.-May 2018, pp. 1967–1970. doi:10.18429/JACoW-IPAC2018-WEPAF062

[5] F. Liu, K.M. Ting, Z. Zhou, "Isolation Forest", in *Proc. 8th IEEE International Conference on Data Mining*, Washington, USA, Dec. 2008, pp. 413–422. doi:10.1109/ICDM.2008.17

[6] A. Langner, G. Benedetti, M. Carlà, U. Iriso, Z. Martí, J. Coello de Portugal, R. Tomás, "Utilizing the N beam position monitor method for turn-by-turn optics measurements", *Phys. Rev. Accel. Beams* vol. 19, p. 092803, 2016. doi:10.1103/PhysRevAccelBeams.19.092803

[7] A. Wegscheider, A. Langner, R. Tomás and A. Franchi, "Analytical N beam position monitor method", *Phys. Rev. Accel. Beams* vol. 20, p. 111002, 2017. doi:10.1103/PhysRevAccelBeams.20.111002

[8] M. Ester, H. Kriegel, X. Sander, J.Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in *Proc. 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96)*, Portland, USA, Aug. 1996, pp. 226–231.

[9] E. Fol, "Detection of faulty Beam Position Monitors", presented at ICFA Beam Dynamics Mini-Workshop: Machine Learning Applications for Particle Accelerators, Menlo Park, CA, USA, 2018.

[10] M. Breunig, H. P. Kriegel, R. T. Ng, J. Sander, "LOF: identifying density-based local outliers", in *ACM sigmod record*, vol. 29 no. 2, pp. 93–104, 2002. doi:10.1145/335191.335388

[11] E. Fol, J. M. Coello de Portugal, and R. Tomás, "Application of Machine Learning to Beam Diagnostics", presented at the 7th International Beam Instrumentation Conference (IBIC'18), Shanghai, China, Sep. 2018, paper TUOA02.

[12] T.G. Dietterich, "Ensemble Methods in Machine Learning", in *Proc. 1st International Workshop on Multiple Classifier Systems*, London, UK, Jun. 2000, pp. 1–15.