ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL03

MACHINE LEARNING BASED MIDDLE-LAYER FOR AUTONOMOUS ACCELERATOR OPERATION AND CONTROL

S. Pioli*, B. Buonomo, F. Cardelli, P. Ciuffetti, L. G. Foggetta, C. Di Giulio,
D. Di Giovenale, G. Piermarini, INFN-LNF, Frascati, Italy
V. Martinelli, INFN-LNL, Legnaro, Italy

Abstract

The Singularity project, led by National Laboratories of Frascati of the National Institute for Nuclear Physics (INFN-LNF), aim to develop automated machine-independent middle-layer to control accelerator operation through machine learning (ML) algorithms like Reinforcement Learning (RL) and Clustering, integrated with accelerator sub-systems. In this work we will present the actual LINAC control system and the necessary effort to implement the architecture and the middle-layer necessary to develop autonomous operation control with RL algorithms together with the fault detection capability improved by Clustering approach as for waveguides or accelerator sections breakdown. Results of the first tentative operation of Singularity on the LINAC system will be reported.

INTRODUCTION

In this paper we will present our effort to integrate a Machine Learning (ML) based middle-layer integrated in the DAΦNE LINAC in order to demonstrate the feasibility of autonomous operation driven by RL algorithm together with Clustering fault detection methods to identify breakdown activities.

The main obstacle to implement the Singularity project [1] in the DAΦNE LINAC is related to the time available for the development and test the system in an operating accelerator for 4500 hours per year. The COVID periods with the reduced activities permits to implement the necessary step to provides at Singularity the data to implements the algorithms.

In the first part of this work introduction of the LINAC elements will be provided and an overview on the the actual control system is shown to introduce the architecture of the system and the middle layer implemented to provides the data to Singularity.

In the second part of the work description of RL algorithms and Singularity middle layer will presented and related integration and performances obtain on off-line operation will be presented and discussed.

THE DAΦNE LINAC

The DAΦNE injector is composed by a ~60 m long Linac that produces and accelerates up to the collider operation energy (510 MeV) both the positron and electron beams. It has been designed and built by the USA firm TITAN BETA and commissioned by the INFN-LNF staff [2]. In Fig. 1 shows

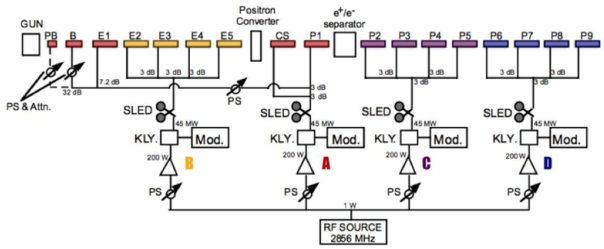

Figure 1: The LINAC layout.

the LINAC RF layout. The injector subsystem includes a thermionic electron gun, a prebuncher and a buncher. The sections performing at S-band working at 2856 MHz, and are powered by 4 klystrons Thomson TH2128C, with nominal output power of 45 MW, each one equipped with a SLED, the SLAC type pulse compressor device. The performances of the LINAC are summarized in Fig. 2.

	Design	Operational
Electron beam final energy	800 MeV	510 MeV
Positron beam final energy	550 MeV	510 MeV
RF frequency	2856 MHz	
Positron conversion energy	250 MeV	220 MeV
Beam pulse rep. rate	1 to 50 Hz	1 to 50 Hz
Beam macropulse length	10 nsec	1.4 to 300 nsec
Gun current	8 A	8 A
Beam spot on positron converter	1 mm	1 mm
norm. Emittance (mm. mrad)	1 (electron) 10 (positron)	< 1.5
rms Energy spread	0.5% (electron) 1.0% (positron)	0.5% (electron) 1.0% (positron)
electron current on positron converter	5 A	5.2 A
Max output electron current	>150 mA	500 mA
Max output positron current	36 mA	85 mA
Trasport efficiency from capture section to linac end	90%	90%
Accelerating structure	SLAC-type, CG, 2π/3	
RF source	4 x 45 MWp sledded klystrons TH2128C	

Figure 2: The LINAC beam performances.

After an upgrade on the gun [3] all the parameters as bunch duration and the other gun parameters and RF power in the RF guide distribution and in the sections could be controlled by klystron voltage set and the low level RF input of each klystron (power and phase) and the prebuncher, buncher, capture section power and phases could be set.

The focusing system varies its conformation according to the requirements of the portion of the LINAC interested.

* stefano.pioli@lnf.infn.it

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL03

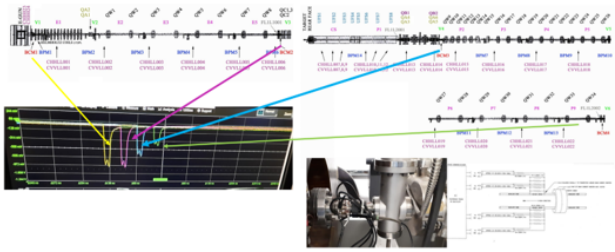

Figure 3: The LINAC magnet elements and diagnostics.

An easy way to describe this system is to follow a particle beam from the gun up to the LINAC end. More than 30 quadrupoles and a network of more than 20 vertical and horizontal correctors, coupled on each of the accelerating section approximately, permits the LINAC orbit correction. All the power supply of this magnets system are controlled by the LINAC control system and part by DAFNE control system.

The LINAC beam diagnostics system which includes a total of 14 beam position monitors BPM, 4 beam current monitors (BCM) and 4 beam profile monitors (flags). The BPMs are acquired by a multiplexed PXI independent system and the WCM by oscilloscope by a HP Multiplexer.

The position monitors are of the capacitive type with four electrodes, see Fig. 3. The voltages induced by the beam on these electrodes, properly combined, give the transverse position of the beam.

The current monitors are of the resistive wall type. The vacuum chamber continuity is interrupted by a ceramic gap and several resistors uniformly distributed around the gap create a resistive bridge, where the beam image current on the vacuum chamber can flow through. By monitoring the voltage drop on the resistors it is possible to derive the current value. Moreover the frequency response of the monitor is sufficient to reproduce the shape of the macrobunch (10 ns) with a good fidelity. Several examples of the current monitor signal are showed in Fig. 3.

To check the energy and the charge of the beam at the and of the LINAC a pulsed magnet drive one of the 50 bunch each second in the hodoscope spectrometer.

THE LINAC CONTROL SYSTEM

A distributed architecture has been chosen for the LINAC control system. These provide the feature to run all or part of the system individually for test and trouble-shooting, and being able to co-locate a subsystem with it's controls.

Originally the control system is operated through a control computer located remotely from the LINAC. An Apple Mac II run National Instruments LabVIEW originally and after a few upgrade in 2013 a server DELL with a virtual machine where the original LabVIEW 3.0 software was upgrade to LabVIEW 2010 on CentOS 7. The control software provides GUI shown in Fig. 4 and the operator full control of all functions and communicate with the CAMAC control system

Figure 4: The LINAC control system GUI.

via an IEEE-488 (GPIB) fiber-optical data link to isolate the computer from the control system.

A remote manual control panel is provided to perform limited hardwire commands and includes hardwire interlock functions. The control computer allow data logging of the system parameters with no impact on system performance. A CAMAC based interface was selected in 1996 for this system on the basis of performance, module availability, and familiarity with programming and operation of CAMAC based systems.

The CAMAC crate and associated controller, GPIB interfaces, and all digital and analog modules are located in the local control rack. The Command and Control buses handles all of the system wide commands and response from all of the subsystems. The Auxiliary buses is used for subsystem specific specific commands and response for the various operational configuration. All of the analog control and monitor signals are not bus-connected, but run direct to/from the required location in the shielded twisted pair cables, using various interface chassis as terminations points. Al the system under control are shown in Fig. 5.

The integration of all the systems and diagnostics are done following the schema represented in Fig. 6 follow the methodology presented in [4] to send and communicate with the Singularity AI. This system allows to integrate different data sources from the LINAC control system with the BPM system, WCM system and DAΦNE control system [5] and to send the command with the same interface.

THE REINFORCEMENT LEARNING APPROACH

According to Sutton's [6], Reinforcement Learning (RL) is a learning technique to map situations onto actions in order to maximize a numerical reward signal. The learner is not told which actions to perform, but must figure out through trial and error which actions yield the greatest reward by trying them. In the most interesting and challenging cases, the actions may affect not only the immediate reward, but also the next situation and through that, all subsequent rewards. These two characteristics (trial-and-error search and delayed

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL03

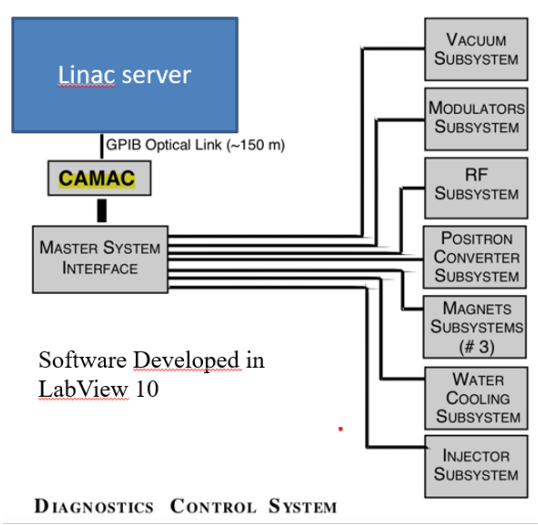

Figure 5: The LINAC control system.

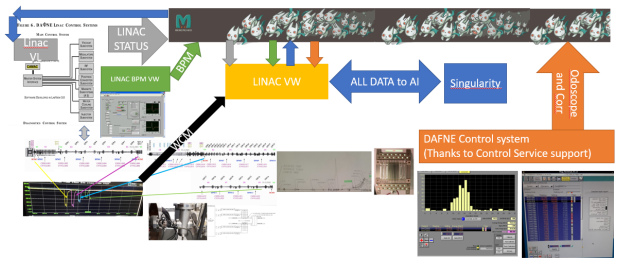

Figure 6: The LINAC control system and diagnostic acquisition.

reward) are the two most important distinguishing features of reinforcement learning.

Reinforcement learning is distinct from supervised learning, the type of learning studied in most current research in machine learning. Supervised learning is learning from a training set of labeled examples provided by a knowledgeable external supervisor. Each example is a description of a situation along with a specification (the label) of the correct action that the system should take in that situation, often to identify a category to which the situation belongs. The goal of this type of learning is for the system to extrapolate or generalize its responses so that it acts correctly in situations not included in the training set.

One of the challenges that reinforcement learning faces is the trade-off between exploration and exploitation that does not exist in other types of learning. To obtain a high reward, a reinforcement learning agent must prefer actions that it has tried in the past and found to be effective in obtaining a reward. However, in order to discover such actions, it must also try actions that it has not previously selected. The agent has to exploit what it has already experienced in order to obtain a reward, but it must also has to explore in order to make a better action selections in the future. The dilemma is that neither exploration nor exploitation can be pursued exclusively without failing at the task. The agent must try a variety of actions and gradually prefer those that seem best to it.

Q-learning [7] is a form of model-free reinforcement learning. It can also be considered as a method of asynchronous dynamic programming. It provides agents with the opportunity to learn to act optimally in Markovian domains, Fig. 7, by experiencing the consequences of their actions without having to create maps of the domains. Learning proceeds similarly to Sutton's [6] method of temporal differences (TD): an agent tries an action in a given state and evaluates its consequences based on the immediate reward or penalty it receives and its estimate of the value of the state it is placed in. By repeatedly trying all actions in all states, it learns which actions are best overall, judged by the long-run discounted reward according to the Bellman transition function.

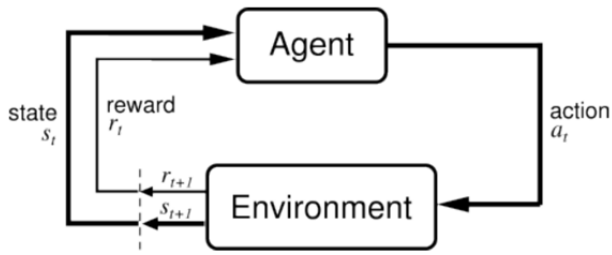

Figure 7: Markov Decision Process.

In this case, the learned action value function Q is a direct approximation to q*, the optimal action value function, regardless of the strategy followed. This drastically simplifies the analysis of the algorithm and allows for early convergence proofs. The policy still has an impact as it determines which state-action pairs are visited and updated. However, all that is required for proper convergence is that all pairs continue to be updated.

Under this assumption and a variant of the usual stochastic approximation conditions for the sequence of step-size parameters, Q converge to q* with probability 1. The Q-learning algorithm is presented below in procedural form in Fig. 8.

CONTROL SYSTEM INTEGRATION & FAULT DETECTION

Mem-cached live data from LINAC control system is acquired from Singularity middle-layer while a REST communication channel allow to change devices setpoint.

In order to allow the proper data validation for ML algorithm, interlock signals are acquired from the CAMAC in order to break and restore operation when a fault happens. Anyway not all the fault trigger an interlock like RF anomaly that should be processed through a data analysis to identify breakdown phenomena and stop the autonomous operation while the issue is occurring.

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL03

Q-learning (off-policy TD control) for estimating $\pi \approx \pi_*$

Algorithm parameters: step size $\alpha \in (0, 1]$, small $\varepsilon > 0$
Initialize $Q(s, a)$, for all $s \in \mathcal{S}^+, a \in \mathcal{A}(s)$, arbitrarily except that $Q(terminal, \cdot) = 0$

Loop for each episode:
  Initialize S
  Loop for each step of episode:
    Choose A from S using policy derived from Q (e.g., ε-greedy)
    Take action A, observe R, S'
    $Q(S, A) \leftarrow Q(S, A) + \alpha\big[R + \gamma \max_a Q(S', a) - Q(S, A)\big]$
    $S \leftarrow S'$
  until S is terminal

Figure 8: Q-learning pseudo-code iteration.

In order to handle this kind of RF events a Clustering algorithm have been involved to automatically recognize pattern inside the data so as to analyze the collected data without their labels. Using this advantage, a density based clustering fault diagnosis method have been used to deal with such RF breakdown issues, in which the labeled data are limited.

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [8] algorithm requires two input parameters, a radius and minimum number used to define a density threshold in the data space. DBSCAN is an iterative algorithm which iterates over the objects in the dataset, analyzing their neighborhood. If there are more than certain number of objects whose distance from the considered object is less than the reference, then the object and its neighborhood originate a new cluster. DBSCAN is effective at finding clusters with arbitrary shape, and it is capable of identifying outliers as a low density area in the data space.

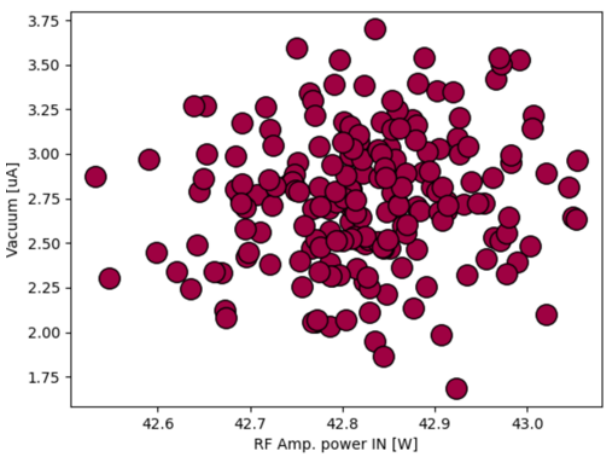

Figure 9: DBSCAN output in normal operation which identify one cluster.

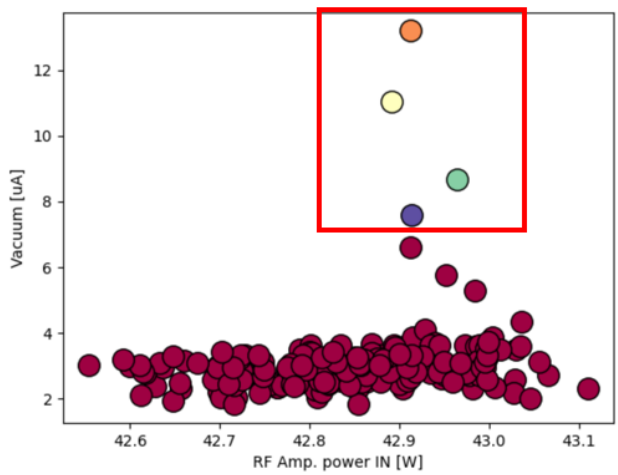

Figure 10: DBSCAN output in breakdown events which identify several clusters.

Thanks to this unsupervised learning technique have been possible control n-dimensional system identified by the evolution of RF power signal and all the vacuum ion pumps current trends in order to identify breakdown. The algorithm is setup to expect one cluster made of a reasonable distribution of point from RF power and ion pumps current distribution, as in Fig. 9. The minimum number of point to identify a cluster is set one in order to allow the classification of fault when just one ion pump show an anomalous behavior. When a vacuum readback moves out from the cluster, as in Fig. 10, this is identified as a second cluster and trigger the identification of a breakdown event stopping the execution of RL algorithm. This process is performed for all the RF sources of the LINAC and all the related vacuum ion pumps.

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL03

AUTONOMOUS OPERATION TOOLS

Two tools for autonomous operation have been developed in order to handle main tasks, really time consuming, which involves the operators team to reach required performances for the DAΦNE complex and the BTF.

As general comment, all the tools have been developed to be capable of machine independence from the accelerator layout and to be invariant to performance drift of aged devices. The combination of these two conditions allow to use these tools independently on different new-one or legacy accelerator just configuring the schematic lattice and operating parameter of each device. The algorithm, during the training phase, will learn how to drive each device identifying for each setpoint the contribute to reach the goal state.

RF Sources Energy Tuning

According with definition of a Markovian Decision Process (MDP), the scope of this tool is the energy tuning of the LINAC beam through the control of two RF Sources (power and phase) after the positron converter and the monitoring the hodoscope at the end of the LINAC which provide the beam energy, as shown in Fig. 11.

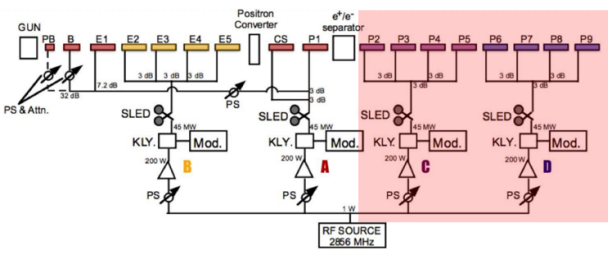

Figure 11: Highlighted in red the RF sources and related distribution controlled by the RF energy tuning tool.

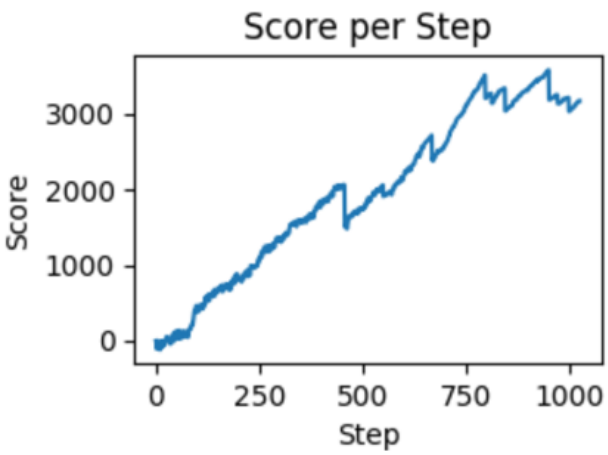

Figure 12: Score trend during each step of the 300 episodes training.

We identify for each variable of the environment operative ranges and the related accuracy. So the state-action matrix comes from the combination of all the possible action of the four variable parameters (power and phase of the two RF sources) with all the possible states (hodoscope energy readback). The Q-learning agent provide variable reward in function of the different between the current beam energy and the target beam energy.

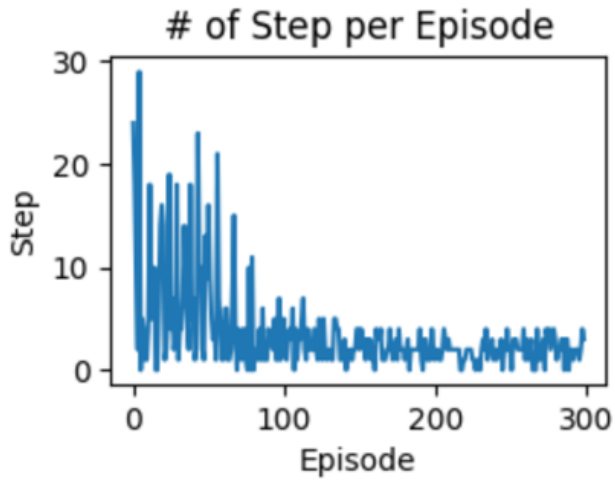

Figure 13: Number of Steps during each one of the 300 episodes training.

A simulated run with off-line data have been performed in order to test the algorithm over 300 episode. As shown in Figs. 12 and 13, the algorithm converge to Q* within 1000 steps equivalent to about one day of training on the LINAC (taking in consideration the required time to perform and stabilize a command with the CAMAC control system).

Beam Charge Optimization

The scope of this tool is optimize the charge of the LINAC beam through the control of 8 quadrupole magnets power supply, after the positron converter, and 2 beam current monitors (BCM) monitoring the beam current up to the end of the LINAC, as shown in Fig. 14.

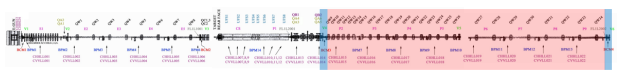

Figure 14: Highlighted in red the distribution of quadrupole magnets while in blue the position of the BCMs controlled by the Beam Current Optimization tool.

We identify for each variable of the environment operative ranges and the related accuracy. So the state-action matrix comes from the combination of all the possible action of the 8 variable parameters (the current setpoint of each power supply) with all the possible states (percentage of current transported through the two BCMs). The Q-learning agent provide variable reward in function of the different between the ratio of current transported and the target beam current ratio.

A simulated run with off-line data have been performed in order to test the algorithm over 300 episode. As shown in Figs. 15 and 16, the algorithm converge to Q* within 2000 steps equivalent to about two days of training on the LINAC

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL03

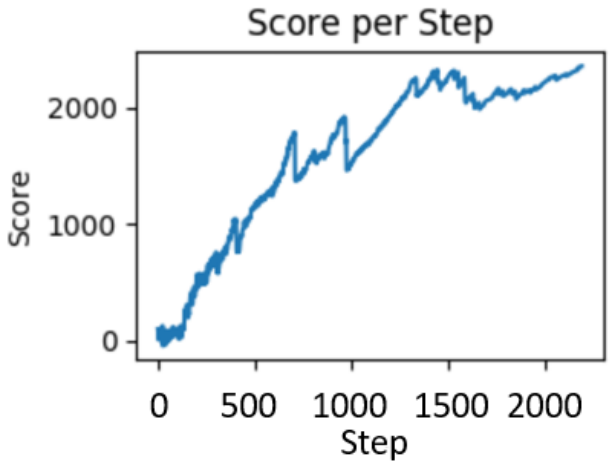

Figure 15: Score trend during each step of the 300 episodes training.

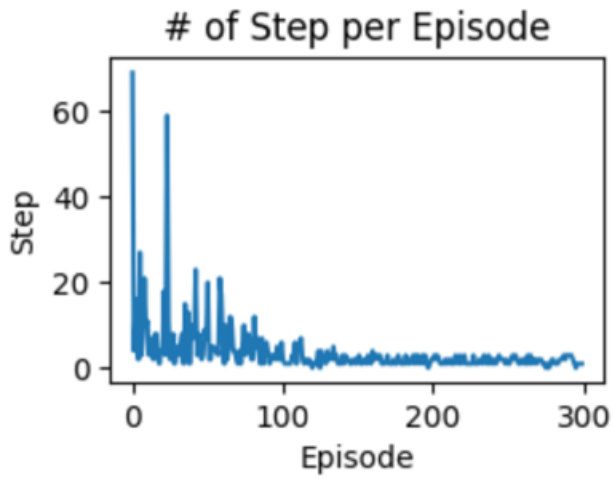

Figure 16: Number of Steps during each one of the 300 episodes training.

(taking in consideration the required time to perform and stabilize a command with the CAMAC control system).

CONCLUSION

The work presented in this paper show the progress in the development of Machine Learning based middle layer focused on Reinforcement Learning tool for automated operations.

Design of the DAΦNE LINAC control system, the first testing site, have been introduced together with the architecture of the Q-learning algorithm.

Validation process on off-line data highlight not only the feasibility of presented tools but also the capability to be trained on a LINAC with a small effort of few days of dedicated beam shift.

Next steps for the use of ML tools in operation will pass through a test shift to be scheduled in next months and an upgrade of the state-action matrix to a Deep Q-Network (DQN) to optimize memory usage.

REFERENCES

[1] S. Pioli, "Expression of Intent Singularity", INFN-19-05/LNF Technical Note

[2] F. Sannibale, M. Vescovi, R. Boni, F. Marcellini, and G. Vignola, "DAΦNE Linac Commissioning Results", DAFNE-NOTE-BM-02, http://www.lnf.infn.it/acceleratori/dafne/NOTEDAFNE/BM/BM-2.pdf

[3] B. Buonomo, L. G. Foggetta, and G. Piermarini, "New Gun Implementation and Performance of the DAΦNE LINAC", in *Proc. IPAC'15*, Richmond, VA, USA, May 2015, pp. 1546–1548, doi:10.18429/JACoW-IPAC2015-TUPWA056

[4] L.G. Foggetta, M. Belli, B. Buonomo, F. Cardelli, R. Ceccarelli, A. Cecchinelli, *et al.*, "The Extended Operative Range of the LNF LINAC and BTF Facilities", in *Proc. IPAC'21*, Campinas, SP, Brazil, May 2021, pp. 3987–3990. doi:10.18429/JACoW-IPAC2021-THPAB113

[5] G. Di Pirro, C. Milardi, A. Stecchi, and L. Trasatti, "DANTE: control system for DAΦNE based on Macintosh and LabVIEW", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 352, pp. 455-457, 1994. doi:10.1016/0168-9002(94)91568-7

[6] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction", 2018, MIT Press, Cambridge, MA, USA.

[7] C. J. C. H. Watkins and P. Dayan, "Q-learning", *Mach Learn*, vol. 8, pp. 279–292, 1992. doi:10.1007/BF00992698

[8] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in *Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD'96)*, pp. 226-231, 1996.