

Pulse Identification Using SVM

Josep PUYOL-GRUART ^{a,1}, Pere GARCIA CALVÉS ^a, Jesús VEGA ^{a,b},
Maria Teresa CEBALLOS ^b, Bea COBO ^b, Francisco J. CARRERA ^b

^aArtificial Intelligence Research Institute (IIIA-CSIC)

^bInstitute of Physics of Cantabria (CSIC-UC)

Abstract.

This paper is about dealing with a noisy signal containing pulses from a detector of X-ray photons. We are designing algorithms to detect pulses, separate them when overlapping, and measure the energy of photons and separation between pulses. In this paper, we present the detection of pulses using SVM.

Keywords. Signal Processing, Machine Learning, Support Vector Machines

1. Introduction

X-ray astronomical observations are performed from space since the Earth's atmosphere blocks this type of radiation. Telemetry limitations to transmit data from the satellites where the X-ray telescopes are installed makes the selection on board of the valuable parts of the signal mandatory. This processing must be done with limited computational resources. We explore machine learning as CNN [1] and SVM to complement current classical methods to improve processing efficiency. As a case study, we use the conditions of the X-IFU [2] detector onboard the ESA Athena mission [3].

2. Experiments

The inputs of our system are noisy digital signals containing single or double pulses, which are the result of X-ray photons impacting Transition Edge Sensor microcalorimeters, the kind of sensor that will be onboard Athena mission. The samples of the signals are separated $6.4\mu s$ (156250 samples/s), and its amplitude is 16 bits data representing the pulse intensity.

We start experimenting with a set of analytically simulated pulses to adjust the training parameters and obtain a first impression of the problem's difficulty. These parameters will be used in the following experiments using a more accurate simulator. This paper is about detecting if there are one or two overlapping pulses in the signal.

¹Corresponding Author: Josep Puyol-Gruart, Artificial Intelligence Research Institute (IIIA-CSIC), Campus UAB zona 2, 08193 Bellaterra, Spain. Email: puyol@iiia.csic.es

2.1. First experiments

We use a very simple simulator to do the first experiments. The following equation can model the typical response of a sensor to a stimulus. We generate signals with one or two pulses and experiment with adding different amounts of Gaussian noise to the signal.

$$I(s) = \frac{E}{\tau_1 - \tau_2} \left(e^{-\frac{1}{\tau_1}(s-s_0)} - e^{-\frac{1}{\tau_2}(s-s_0)} \right)$$

$I(s)$ is the pulse intensity at sample s , with constants: E the energy of the pulse, sample constants τ_1 and τ_2 configure the signal's shape, and s_0 is the sample origin. In these experiments $\tau_1 = 30$ and $\tau_2 = 50$ produce shapes similar to the real ones.

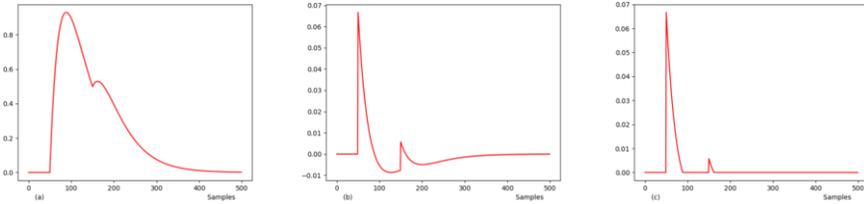


Figure 1. (a) Two pulses starting at samples 50 and 150, with energies 100 and 20 respectively. (b) First derivative (c) First derivative without negative values.

We experimented with the standard signal, its first derivative and cutting the first derivative's negative part (see Figure 1). In all cases, we generate training and test sets of different sizes: 10K, 25K, 50K, 75K and 100K examples with the same number of simple and double pulses. We use 65536 points of amplitude (corresponding to 16 bits of information) and a window of 500 samples. After different experiments with libSVM[4], we use C-SVC, and a kernel with a radial basis function, with $cost = 4$ and the default $\gamma = 1/500$. Rescaling seems not adequate to this problem giving worse results.

In these first experiments, we obtained excellent results, with accuracy greater than 98% with 50K training examples. Never a single pulse is seen as double, but sometimes a double pulse is considered to be single. As we can expect, we obtain the model using the best results with the positive part of the first derivative.

It is essential to notice that preprocessing does not add computational cost to the process. Simple hardware can make this kind of operation before the digitalisation of the signal.

2.2. Experiments with SIXTE simulations

The preliminary experiments described in the section above give us a good idea of the parameters that can be useful to obtain an accurate model with more realistic inputs. Now we will use the same criteria to apply our methodology to a more detailed simulator used for several X-ray missions. SIXTE [5] environment contains the Athena official software for simulations.

We use 60K files containing the examples generated by SIXTE, the same used in the CNN experiments in [1]. We use 45K examples for training and 15K for testing (7500

singles and 7500 doubles). We use the same parameters as in the section before. The input signal is the first derivative without negative elements (*IDer-Cut*).

Actual		Predicted					
		<i>IDer-Cut</i>		<i>IDer-Cut-Reverse</i>		<i>2models</i>	
		One	Two	One	Two	One	Two
One	7500	0	7499	1	7499	1	
Two	291	7209	96	7404	40	7460	

Table 1. Confusion matrices for the models *IDer-Cut*, *IDer-Cut-Reverse* and the combination of both.

These first experimental results (see Table 1 in column *IDer-Cut*) are similar to those obtained with the first simple simulator. There are no false double pulses, but there are 291 false single pulses (3.88%). All pulses classified by the model as double are true doubles.

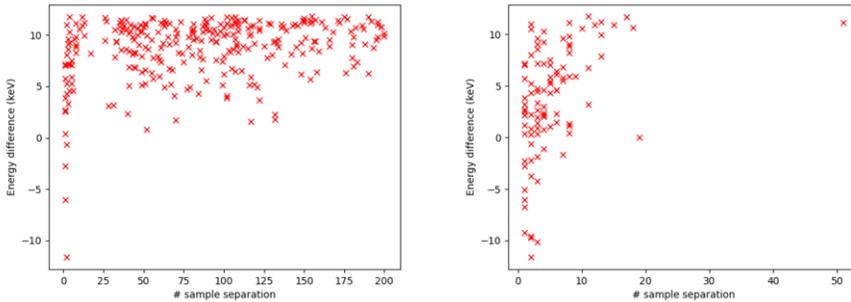


Figure 2. Double pulses detected as singles for the models *IDer-Cut* (left) and *IDer-Cut-Reverse* (right). The X-axis shows the separation of the pulses starting points, and the Y-axis shows their difference of energies.

The separation in time and the difference of energies between two pulses affects the possibility of detecting double pulses. The errors in detecting double pulses are concentrated in pulses separated by relatively few samples or with a significant positive difference of energies, as we can see in Figure 2 left. These errors seem apparent, but we can observe that it is more challenging to detect two pulses when the first pulse has more energy than the second. On the other hand, there are no errors when the first pulse is smaller than the second, except when a few samples separate them. This fact suggests doing a new experiment reading the signals in reverse and obtaining a new model (*IDer-Cut-Reverse*). We convert double pulses with the second smaller to double pulses with the first smaller. The results are in Table 1 in column *IDer-Cut-Reverse* and Figure 2 right. The results are better than in the model from the non reversed signal, with only 96 false single pulses (1.28%).

Finally, we decide to combine the two models (*IDer-Cut* and *IDer-Cut-Reverse*) to deduce if the signal contains single or double pulses. We use the two models and, when both models consider that the pulse is single, then the result is a single pulse. Otherwise, the result is a double pulse.

We can see in Table 1 in column *2models* that the false single pulses are 40 (0.53%) and that now appears one false double pulse. This false double pulse corresponds to a

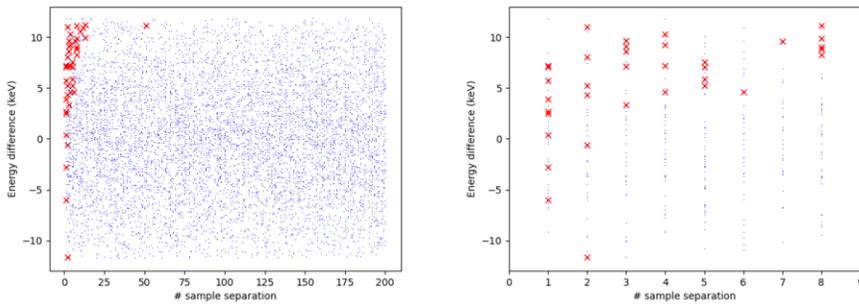


Figure 3. Representation of double pulses detected as singles for the model *2models* (red crosses) and double pulses correctly detected (blue points) At the right, a zoomed version of the left.

single pulse with an energy of 11896eV. It is close to the limit of saturation of the detector (12keV), which can be the reason for the malfunction. The deformation of the signal's shape confuses the model and detects a double pulse.

In Figure 3, we can see at left the distribution of false singles by separation in time and difference of energies. Notice that using the two models, false singles are produced clearly in close pulses or considerable difference of energies, especially when the first pulse is more energetic than the second.

There is an exception in the lonely red cross. This false single pulse corresponds to a double pulse with energies 11319eV and 0.205eV separated by 51 samples. The big difference in energies and the energy of the second pulse (at the limit of detection, 0.2eV) can be the reason for this outlier.

Figure 3 right is a zoom of the left one representing pulses with separations of less than ten samples. We can see that there are nine examples of double pulses only separated by one sample and six by two. When the separation increases, the difference of energies becomes more critical when the first pulse has more energy than the second. The results are very satisfactory, with the model only failing in a small fraction of extreme cases.

3. Conclusions and Future Work

These experiments demonstrate that good detection of single and double pulses is possible using SVM. Although the results are close, an estimate of the computational cost of SVM should be made and compared to CNN [1]. Currently, we are working on estimating the separation between pulses with good prospects. We will also attempt to estimate the pulses' energy, although with less hope since the good results of classical techniques seem challenging to surpass.

References

- [1] J. Vega et al., 2020. TES X-ray pulse identification using CNNs. ADASS XXX.
- [2] Barret D., et al., 2018, in SPIE. p. 106991G (arXiv:1807.06092),doi:10.1117/12.2312409.
- [3] Nandra K., et al., 2013, arXiv e-prints, p. arXiv:1306.2307.
- [4] C.-C. Chang and C.-J. Lin, 2011. LIBSVM: A library for support vector machines. TIST, 2(3):127.
- [5] Dauser et al. 2019, A&A, 630, A66.