30TH INTERNATIONAL COSMIC RAY CONFERENCE

Classification methods for Cherenkov telescopes images on a pixel-by-pixel base

C. MALAGON [1], J.A. BARRIO[2], D. NIETO [2], R. DE LOS REYES [2]
[1]*Dpto. Ingeniería Informatica, Univ. Antonio de Nebrija, 28040 Madrid, Spain.*
[2]*Dpto. Física Atomica, Univ. Complutense, 28040 Madrid, Spain*
cmalagon@nebrija.es, barrio@gae.ucm.es

Abstract: The problem of identifying gamma ray events out of charged cosmic ray background (so called hadrons) in Cherenkov telescopes is one of the key problems in VHE gamma ray astronomy. In this contribution, we present a novel approach to this problem by implementing different classifiers relying on the information of each pixel of the camera of a Cherenkov telescope, rather than using common Hillas parameter analysis. Separation between gamma-like and hadron-like is performed using several machine learning techniques, trained using Monte Carlo data samples of both types of events.

Introduction

Since Hillas parameter analysis was developed back in 1985 to separate between gamma-like and hadron-like events as recorded by Cherenkov telescopes [1], many techniques have been used for gamma/hadron separation based on such parameters. Bock et al. [2] performed a case study for most of these techniques, to be later applied to MAGIC telescope gamma event selection. However, all these techniques might not be using the whole potential of a Cherenkov telescope, as they use Hillas parameters (second moments of image in telescope camera) as input. In this work we propose to apply usual machine learning techniques (some of them, mentioned in [2]) to the full image recorded by a Cherenkov telescope, on a pixel-by-pixel base. We will demonstrate the method on images produced by the MAGIC telescope simulation and reconstruction package [3, 4]. Possible advantages of this approach are the use of the full information in the camera, and a more natural way to treat fluctuations in the image, thus permitting a relaxation of the image cleaning and a likely reduction of the telescope software threshold. Some inconvenients of this approach come from the fact that certain effects very difficult to simulate (like bright stars in the Field of View), can play a certain role in a pixel-by-pixel analysis, while only a minor effect on a Hillas parameter analysis. Another approach also based on pixel-by-pixel information was already developed by Le Bohec et al. [5] for the CAT telescope. However, this approach did not use machine learning techniques, but a maximum likelihood technique to compare images from simulated showers against analytical expressions from shower development in the atmosphere.

Data sample

For this study, we have used gamma and proton events (as the latter represents the majority of hadronic cosmic rays) simulated with Corsika code [6], plus MAGIC Reflector and Camera reconstruction standard software. Each event consists of an image based on the calibrated photoelectron content in each pixel of the MAGIC telescope camera. The pixels whose signal is likely to originate from NSB or electronic noise are removed from the image using the so-called image cleaning procedure [7], with 10 photoelectron threshold for core pixels and 5 photoelectron threshold for boundary pixels.

Gamma and proton samples consist of 28750 events. Image total photoelectron spectrum of gamma sample resembles that of typical cosmic sources. Corresponding spectrum of proton sample is forced with similar slope to avoid biasing the selection procedure.

In what follows, we will briefly describe the different classifiers used in the experiment, composed by individual or ensemble classifiers.

To begin with, we have used three individualclassifiers for our experiments: A brief description of each of them follows, with a deeper description in references below:

- Decision trees - construction of decision or classification trees using Quinlan's C4.5 algorithm [8].

- Multilayer Perceptron - a feed forward artificial neural network trained with the classical backpropagation algorithm [9] .

- K-NN - K Nearest Neighbour algorithm [10] with K=11 and euclidean distance metric.

Regarding ensemble classifiers, which will only be applied to decision trees classifier, we will deal with multiple classifier systems generally described as voting classification algorithms, i.e., techniques in which we use several individual classifiers that output a particular prediction or label for each of the examples of the test data set. These predictions are then combined to produce a single output, the output of the ensemble, by majority voting decision.

We used two different voting algorithms:

- Boosting - We used the Adaboost algorithm [11] implemented in Weka [12].

- Bagging - Classical ensemble method developed by Breiman [13].

Let us make a brief description of this two ensemble methods: both are combinations of individual classifier (decision trees in our case) and they output a prediction based on majorityvote. The main difference between these algorithms is the way they build the training set. Bagging construct training subsets sampling from the original training set with replacement (i.e. some examples could be repeated in each subset). Then it builds a model for each of the subsets and combines the output of the different models typically by majority voting. Boosting method builds up the different subsets by samplingexamples without replacements. The key point is that boosting method constructs every model paying special attention to those examples that previous model classified incorrectly.

Results

All these machine learning methods have been fed with above described gamma and proton samples, using a typical holdout validation method (two thirds of them for training and the rest for testing purposes).

Results from the different methods, presented in terms of ROCc (Receiver Operating Characteristic curves) for different classification methods, are shown in figures bellow: vertical axis shows gamma acceptance while horizontal one represents hadron acceptance. Q (quality factor) curve for energies lower than 200 GeV is also presented; Q curve is defined as $Q = \epsilon_\gamma / \sqrt{\epsilon_h}$, where ϵ_γ is the gamma acceptance and ϵ_h is the hadron acceptance.

All the methods have been applied both to the whole gamma and proton sample and to subsamples, depending of incident energies. Figure 1 and 2 show the ROC and Q factor curves respectively for samples with energies lower than 200 GeV. Both where compared with results in López [14], graphically displayed using point style. López results represent a practical realization of Bock et al. [2] *Random Forest* classification method (equivalent to out *Bagging* ensemble of decision trees), applied to MAGIC telescope current configuration.
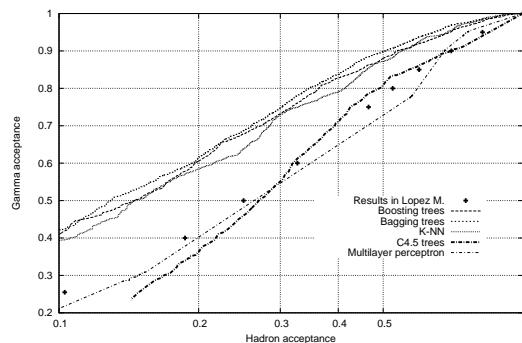

Figure 1: ROCc for energies lower than 200 GeV

One has to emphasize that all methods are applied on a pixel by-pixel basis, even classification trees ones. Ensemble methods (also known as multiple classifier systems) show superior behaviour as compared to its individual classifier (decision trees), and similar to K-NN method. Only within highest energies decision trees (the weak learner we have combined in ensembles) perform well. On the other hand, artificial neural network shows the worst performance of the whole set of classifiers.
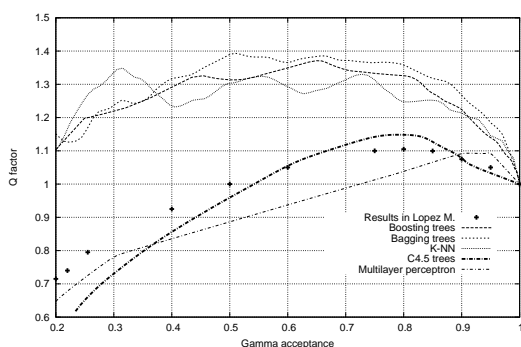

Figure 2: Q factor vs hadron acceptance for energies lower than 200 GeV

Figure 3 shows the results when classifying the whole data set. Boosting trees show results comparable with those in Bock et al [2], graphically displayed using point style as well.
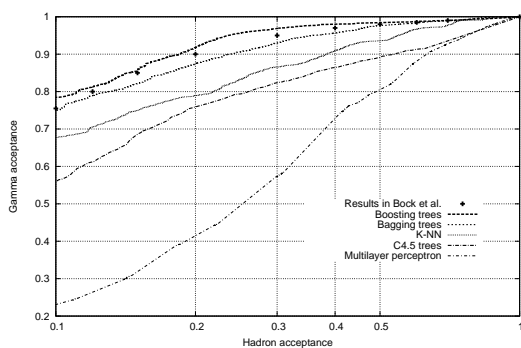

Figure 3: ROCc for total events

Discussion of results

Let us first examine the results of Nearest Neighbours algorithm. Our first intuition, based on do-

main knowledge when we began this work, was that Nearest Neighbours could be a good choice. This intuition was based on the fact that K-NN has a strong geometrical character.

Nearness is calculated using a similarity measure between the different instances, typically the euclidean distance. Therefore it has a strong geometrical dependence, as our domain representation, and therefore we can include a priori information on the camera geometry. We believe this is the reason why Nearest Neighbours technique could stand out from the other single predictors, and results have confirmed our early prediction.

Regarding the ensemble methods, we have combined decision trees, and we have obtained comparable results to those from Hillas parameter analysis in previous works [2]. Decision trees is a good model to be combined in ensembles because of its large variance (different training sets lead to completely different models). These combinations normally result in an enhancing of the accuracy of any single decision tree, though it depends on the domain characteristics.

The behaviour for events with energies lower than 200 Gev show a sizeable improvement with respect to results in López [14], both for ROC and Q curves. Q for a gamma acceptance level of 0.7 (typical value for MAGIC standard analysis) shows an increase of 25%. In terms of the ROC curve, if we consider for example a value of 0.2 for hadron acceptance, we obtain an increase close to 50% for gamma acceptance.

Conclusions and outlook

Separation between simulated gamma-like and hadron-like events (as reconstructed by the MAGIC Cherenkov Telescope) is performed using several machine learning techniques applied to pixel-by-pixel defined images. Both ensembles of classification trees and K Nearest Neighbours show similar performance as for Bock et al. [2] classifications methods (without any restriction on the energy of the primaries). Our results also show a better performance as compared to those in López [14] for events with energies lower than 200 GeV.

Among all individual classifiers, the algorithm that shows the best accuracy is K Nearest Neighbours.

This is a very interesting algorithm (regarding our domain representation) that shows a similar result than multiples classification systems at lower energies and outperform them when higher energies are also considered. This is a consequence of the strong geometrical dependence of this domain as we are solving the problem of identify gamma ray events relying on the information of each pixel of the camera of a Cherenkov telescope. These classifiers were also discussed in terms of event energies, showing promising results for events with energies below 200 GeV. We have demonstrated that this approach could be comparable to common Hillas parameter analysis but without the requirement of any additional data transformation.

Acknowledgements

We want to thank the MAGIC collaboration for kindly providing us with the simulated gamma and proton sample that we have used to test the method. We also want to thank R. Bock for fruitful discussions.

References

[1] Hillas A. M.: Cherenkov light emission of extensive air showers produced by primary ? rays and by nuclei, Proc. 19th ICRC, La Jolla, 4, 445 ,1985.

[2] Bock, R.K. et al.: Methods for multidimensional event classification: a case study using images from a Cherenkov gamma-ray telescope, Nucl. Instrum. MethA516, 511-528 , 2004.

[3] Bretz, T. (MAGIC Coll.): Procs. 29th ICRC, Pune, India, 2005.

[4] Gaug, M., et al (MAGIC Coll.): Procs. 29th ICRC, Pune, India, 2005.

[5] Le Bohec, S. et al.: A new analysis method for very high definition Imaging Atmospheric Cherenkov Telescopes as applied to the CAT telescope. Nucl.Instrum.Meth. A416, 425-437,1998.

[6] D. Heck et al., CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers

[7] D.J.Fegan: Gamma/hadron separation at TeV energies, Topical Review, J.Phys.G: Nucl. Part. Phys. 23, 1013, 1997.

[8] Quinlan, R.: Induction of Decision Trees. Machine Learning,1, 81-106, 1986.

[9] Rumelhart D., McClelland J. And the PDP Research Group. Learning internal representations by error propagation. Pararell Distributed processing: Explorations in the microstructure of cognition. 1-3. MIT Press, 1986.

[10] Aha, D., and Kibler, D.: Instance-based learning algorithms. Machine Learning, 6, 37-66, 1991.

[11] Freund, Y., Schapire, R. E.: Experiments with a new boosting algorithm, In Proc. 13th International Conference on Machine Learning. San Francisco: Morgan Kaufmann. 148-146, 1996.

[12] Witten, Ian H., Eibe Frank.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations Morgan Kaufmann, San Francisco, 2000.

[13] Breiman, L.: Bagging predictors. Machine Learning 24, 123-140, 1996.

[14] López M., PhD. thesis, UCM, Madrid 2006