Using machine learning techniques for Gamma/Hadron separation in HAWC

Name:Fan Kwok Lung
University No.:3035272731

Fan Kwok Lung

Internal supervisor:Dr. Stephen Chi Yung Ng
External supervisor: Prof. Jordan Goodman,University of Maryland

Introduction

High-Altitude Water Cherenkov (HAWC) Observatory is a ground-based water Cherenkov telescope which detects Very High Energy gamma rays indirect. It detects the Cherenkov radiation emitted when the secondary products of high energy photons hit the water in the tanks and then reconstruct the original event[1]. The study of very high energy (VHE, >100 GeV) gamma rays can lead to a new understanding of energetic astrophysical processes such as dark matter annihilation, active galactic nuclei, and pulsar acceleration mechanisms. However, besides such VHE photons, high energy cosmic rays also interact with the atmosphere and produce similar secondary products. Since HAWC triggers 20000 events per second, a method that is able to separate the cosmic rays from the gamma rays efficiently and reliably is needed.

Figure 1. The High-Altitude Water Cherenkov (HAWC) Observatory

Methodology

When HAWC array detects an event, it will run some events reconstruction algorithm to reconstruct the event and output the reconstruction variables. For example, the core of the air shower which is fitted by an NKG function with maximum likelihood and the energy of the events.Some reconstructed variables will have the power of predicting whether the event is a hadron or a photon. In the view of the physical process, hadron will generate a lot of muons and photons will not. Therefore, a hadron event will usually have some clump far away from the air shower core. As a result, the chi-square error of fitting the lateral distribution to hadron event will usually be higher. Also, other variables might hide the similar information and help to predict the nature of the events.

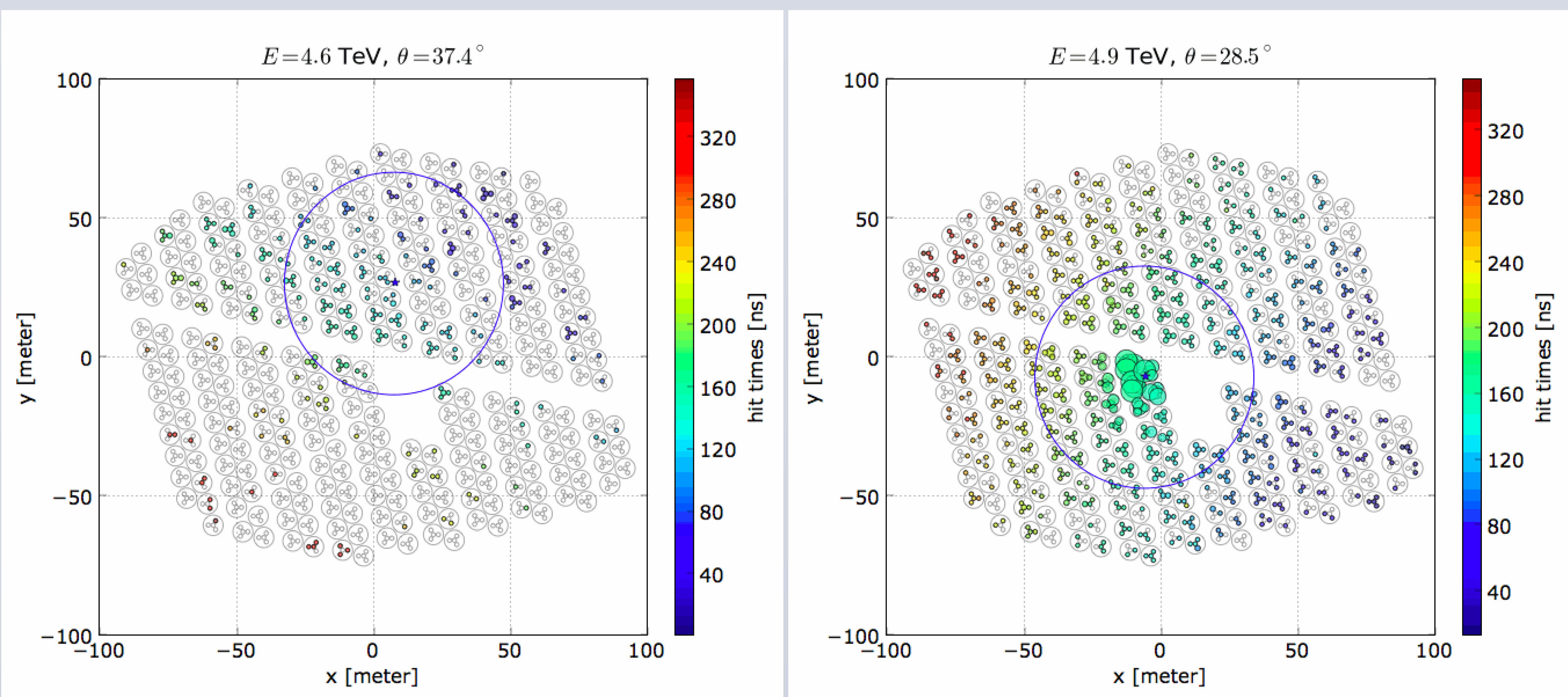

Figure 2. The left-hand side is an example of hadron event. The right-hand side is an example of gamma event.

Methodology(Cont'd)

Gradient boosting decision trees is an extremely powerful tool for such a two-class classification problem and it is the primary algorithms used in this project. Since real gamma rays data is not available, simulated gamma-ray and hadron data are used in the modeling training and testing process. The simulation takes every physical process and detector response into account and then compare with the real data. We will use 50% data to train the model and 50% to test the model. Since the behavior of the events changes with fraction hit of the detector and the energy. We divide the events into 10 fractions hit bins and 12 energy bins. In each bin, we iterator over different probability threshold to obtain a threshold that maximizes q factor= $\frac{True\ positive\ rate}{\sqrt{False\ positve\ rate}} \propto \frac{signal}{\sqrt{background}}$[2]. Since the background signal should follow a Poisson distribution, this value is essentially the Poisson significance using normal distribution approximation.

Result

The current Gamma/Hadron Separation method used is made hard cut in the reconstruction variables. By comparing the q factor of the model and the hard cut with the simulated data, we found that the q factor of the model is higher than the hard cut, especially in the higher energy and intermediate fraction hit bins.

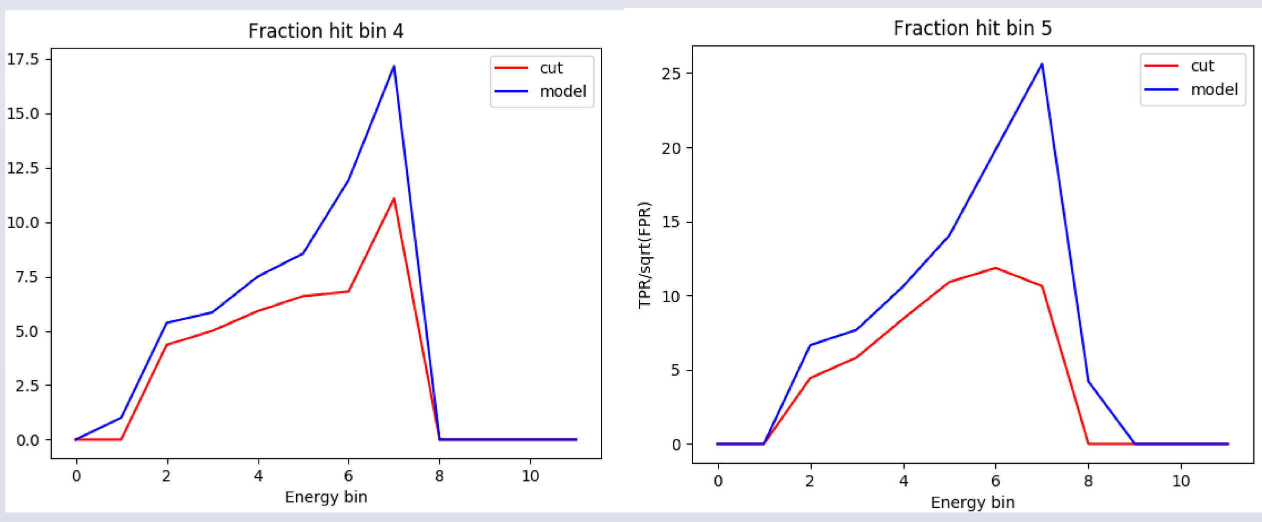

Figure 3. The q factor comparison between the model(blue) and hard cut(red) in fraction hit bin 4 and 5. The x axis is the energy bin and y axis is the q factor. The model have a higher q factor than the hard cut in all the bins.

However, using simulated data to test the performance is not rigorous enough. We apply the method on real data and evaluate the significance on a known bright gamma-ray source. Crab pulsar in the brightest gamma-ray source, we evaluate the significance of it in different bins using two years of real data. The model results in a higher significance in most of the bins and the method is further verified.

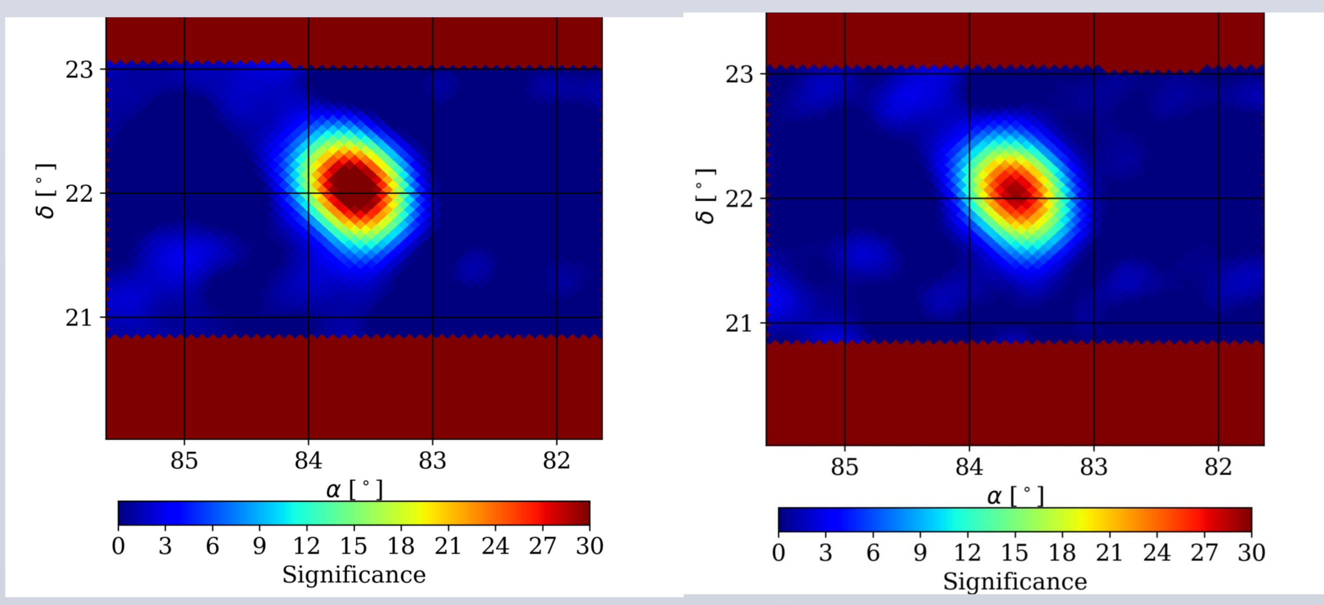

Figure 4. The left hand side the crab significance map using the model and the right is the hard cut in one bin. Using the model result in a higher significance.

References

[1] Abeysekara, A., Alfaro, R., Alvarez, C., Á lvarez, J., Arceo, R., ArteagaVelázquez, J., Solares, H. A., Barber, A., Baughman, B., Bautista-Elivar, N., et al. Sensitivity of the high altitude water cherenkov detector to sources of multi-tev gamma rays. Astroparticle Physics 50 (2013), 26–32.

[2] Capistrán, Tomás, I. Torres, and L. Altamirano. "New method for Gamma/Hadron separation in HAWC using neural networks." *arXiv preprint arXiv:1508.04370* (2015).