

A Multivariate Approach for Identification of Light Nuclear Isotopes of the Cosmic-Rays with Magnetic Spectrometers

Wooyoung Jang¹

*Kyungpook National University
702-701 Daegu, Korea
E-mail: wyjang@knu.ac.kr*

Dongchul Son

*Kyungpook National University
702-701 Daegu, Korea
E-mail: son@knu.ac.kr*

Chanhoon Chung

*I. Physikalisches Institut B, RWTH
D-52056 Aachen, Germany
E-mail: chchung@rwth-aachen.de*

In this work, we will show a way to distinguish various kinds of light nuclear isotopes in cosmic-rays using boosted decision tree. Magnetic spectrometers are typically equipped with a magnet, a tracking system and a time-of-flight. This combination of apparatus can be used to reconstruct mass information of incoming charged particles from measurements of rigidity, speed and charge so that we can identify the incoming particles by their masses. A multivariate approach is found to be helpful to determine the separation power of particle identification among light isotopes.

*35th International Cosmic Ray Conference - ICRC2017
10-20 July, 2017
Bexco, Busan, Korea*

¹Speaker

1. Introduction

Contemporary detector systems for high energy astroparticle physics research usually determine physical variables with multiple simultaneous measurements. This is usually originated from the complexity of experimental goal or the result of pursuit of generality of the detector or to improve precision of measurement. This environment gives us chances to apply multivariate analysis method and may it can be possible to improve our analysis result compare to the classical approaches in some cases. From the past few decades, physicists actively adopted multivariate analysis methods of determining variables for better selection of signals out of backgrounds.

In this contribution, we will report our study about the application of multivariate approach, especially boosted decision tree(BDT) algorithm, to distinguish various kinds of light nuclei of cosmic-rays.

2. Methods

2.1 Benchmarks of Previous Experiments and Construction of Toy MC

We bench-marked the well known magnetic spectrometers such as AMS-02, PAMELA and so on. These magnetic spectrometers are different in details, but in many points they share common features.

Since those two detectors are magnetic spectrometers which has permanent magnets to distinguish sign of charge of incoming particle. AMS-02 has Nd-Fe-B ferromagnetic magnet which has field intensity 0.15 T at the center of the magnet cavity[1]. PAMELA also has Nd-Fe-B permanent magnet which has 0.48 T at the center of the magnet cavity[2].

To determine speed of incident particle, we need Time-of-Flight(TOF) system. In many cases, Time-of-Flight system consist of scintillation counters at the top of the magnet and at the bottom of the magnet. The AMS-02 TOF system comprises 4 planes of polyvinyltoluene scintillation counters, two of them located on the top of the magnet and the other two of them located at the bottom of the magnet. Each 4 alternate layers are placed perpendicularly to each other so that they can be used to record coordinate information. The AMS-02 TOF system can measure the timing information of the incoming particle with precision up to 150 ps. This timing resolution corresponds to approximately 1% of uncertainty of speed measurement when the rigidity of incoming particle is less than 10 GV/c.[3] The PAMELA TOF system consist of six layers of fast plastic scintillation counters arranged in three planes. Those six layers are also placed perpendicularly against the alternating layers for coordinate measurement. The PAMELA TOF system showed it's in-orbit performance of beta measurement around 1.3% for proton events with rigidity $R > 12$ GV/c and the timing resolution is found to be approximately ~100 ps.[4]

Inside the magnet, tracking systems are placed. These days, magnetic spectrometers are usually adopted silicon sensors for precision measurement of particle position. The AMS-02 Tracker consist of 9 layers of double-sided silicon micro-strip sensors. Three layers out of nine are located outside of the magnet and the others are placed inside of the magnet. Position resolution of the AMS-02 Tracker is found to be 10~20um in the bending plane and rigidity

resolution against 10 GV proton is approximately 10%. [1] The PAMELA Tracker is also made of silicon sensors and six layers are located inside the magnet. Position resolution is around 10um for bending plane measurement and rigidity resolution is around 10% for proton events with rigidity less than 10 GV/c. [5]

Based on those similarity between both two standard magnetic spectrometers, we built a simple Monte Carlo geometry using GEANT4.10.02 simulation toolkit[6]. This toy MC detector has 1m height cylindrical magnetic field with 1m diameter. The field intensity is assumed to be 0.15 T.

MC TOF system is consist of four layers of polyvinyltoluene boxes which covers the top and bottom of the magnet. Each layer has 1cm thickness and it's dimension is 1m x 1m in x-y plane. We calculated the timing difference between upper layer and lower layer based on pass-length of the track and momentum of incident particle, after then we applied the 150ps timing resolution effect by adding random Gaussian number to the calculated timing difference.

MC Tracking system is made by six layers of SiO₂ material with 300um thickness and we applied 10um uncertainty for position measurements of each layers. With respect to this setting, the MC tracker has approximately 10% momentum resolution for proton events with rigidity less than 10GV/c. The characteristics of two real detectors, AMS-02 and PAMELA and of the MC detector is summarized at the Table 1.

Table 1 Summary and comparison of characteristics of well-known space-borne magnetic spectrometers with the MC detector features constructed for this study.

	AMS-02	PAMELA	MC
Timing res.	150 ps	100~200ps	150 ps
Beta res.	~1% (R<10GV/c)	~1.3% (R<10GV/c)	1% (R<10GV/c)
Position res. (bending)	10~20um	10~20um	10um
Rigidity res.	~10%	~10%	10%

2.2 Classification of Events using Boosted Decision Tree

Boosted decision tree is one of the algorithms provided by the multivariate analysis (MVA) package TMVA[7]. In this study, we used TMVA version 4.2.1. Multivariate analysis is usually regarded as an extensions of simple cut-based analysis and this techniques are began to use widely in natural scientists to classify various types of events which can be described by multiple number of parameters. After the successful implementation and application of the BDT method at the MiniBooNE [8] neutrino experiment, lots of people in high energy physics adopted this method for their own purposes, for example, AMS-02 experiment also utilized this method to distinguish electrons/positrons from the proton backgrounds using their Electromagnetic Calorimeter(ECAL)[9].

TMVA provides several different types of MVA methods besides of BDT, such as, Likelihood Estimators, Artificial Neural Networks(ANNs) or Fisher discriminants. The BDT is mainly differentiated by it's consideration of non-linear correlations between input parameters.

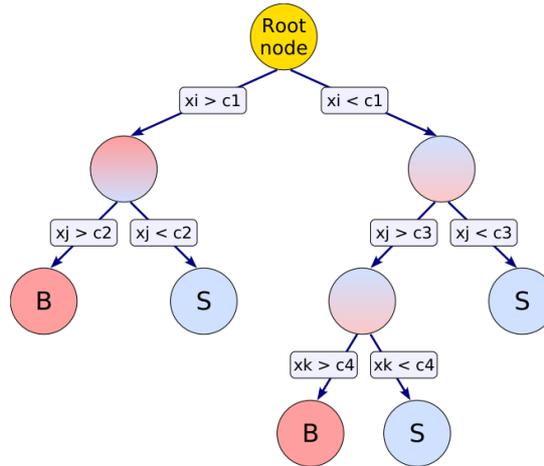


Fig. 1 Schematic diagram of a decision tree. Each of events can be described by a set of parameter x_i , pass through each node a binary split criterion is applied until it ends up in a leaf. At this leaf we mark it as signal(S) or background(B)

2.2.1 Basic Knowledge about the Decision Tree

Decision tree[10] can be represented as a subsequent series of binary split structure like the drawing sketched on the Fig. 1. Applying a binary split criterion at the each of branching, one of the characterizing input parameters used to classify events of unknown type as signal-like or background-like. Determining those each of criteria is one of the main routines of the *training* process of a decision tree. This training process is performed with a training sample which is composed of events of known type whether each of these is signal or background. One of the well-known problems of decision tree algorithm is the instability against statistical fluctuations in the training sample. To resolve this problem one extends the single decision tree to a *forest* of decision trees. The forest is comprises different shapes of decision trees that each of binary splitting is placed differently to the other trees. For each of single decision trees in the forest a weighted mean vote stabilizes the response of the classifier and improves its performance. We use the result of voting as the output of the BDT and it describes the signal- or background-likelihood of an event.

2.2.2 Training parameters

So far, we have discussed the fundamental knowledge of BDT, and now, in this section we will talk about the training and evaluation of the BDT method. We introduced the training parameters used for this work and the distinctive characteristics of the signal- and background training sample. After then, we presented the result of test of the classifier response.

When the incident particle hit all the trackers and TOF planes, since we have six layers of silicon trackers and four planes of TOF scintillators, we could obtain 11 different energy deposition information per event.

Since the deposited energy is depends on the material of the detector and characteristics of incident particle, signal events and background events leave different shape of energy distribution. According to the theoretical works, energy loss rate of particle that traversing certain medium can be estimated by theoretical calculation known as Bethe formula, [11]

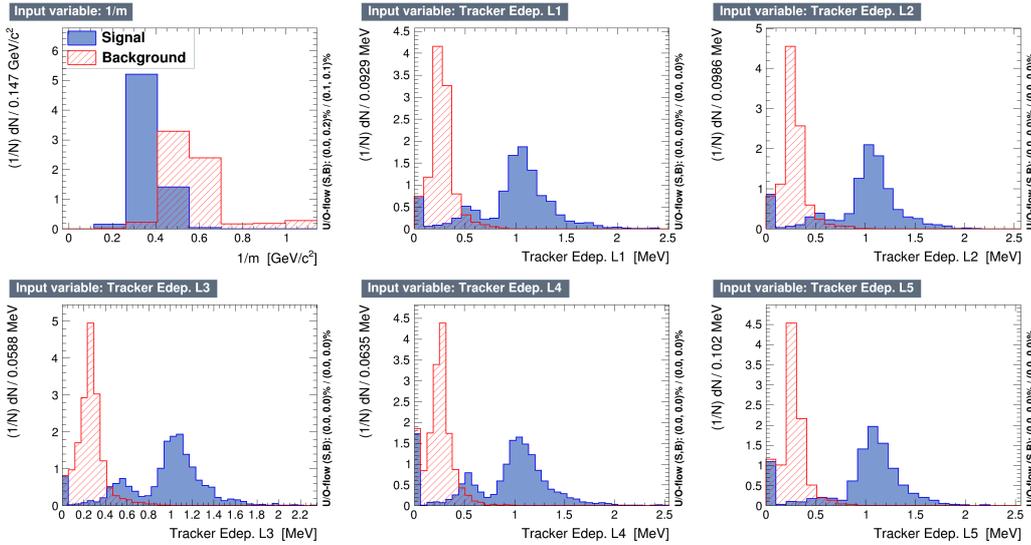


Fig. 2 Example of distribution of the training variables in a fixed beta between (0.7~0.71) for deuteron(blue) and helium-3 (red). Upper left-most is mass distribution and the others for energy deposition on each tracker and TOF layers.

If we are interested about distinguishing different types of nuclear species, due to energy loss rate is proportional to square of charge of incident particle, the charge itself will be a dominating factor to distinguish the energy spectra of different nuclear species. Since we have 11 chances to measure energy deposition, we expected using multivariate analysis will help us to improve the separation of those differently charged particles. Fig. 2 shows the distribution of training parameters of signal sample and background sample for this case.

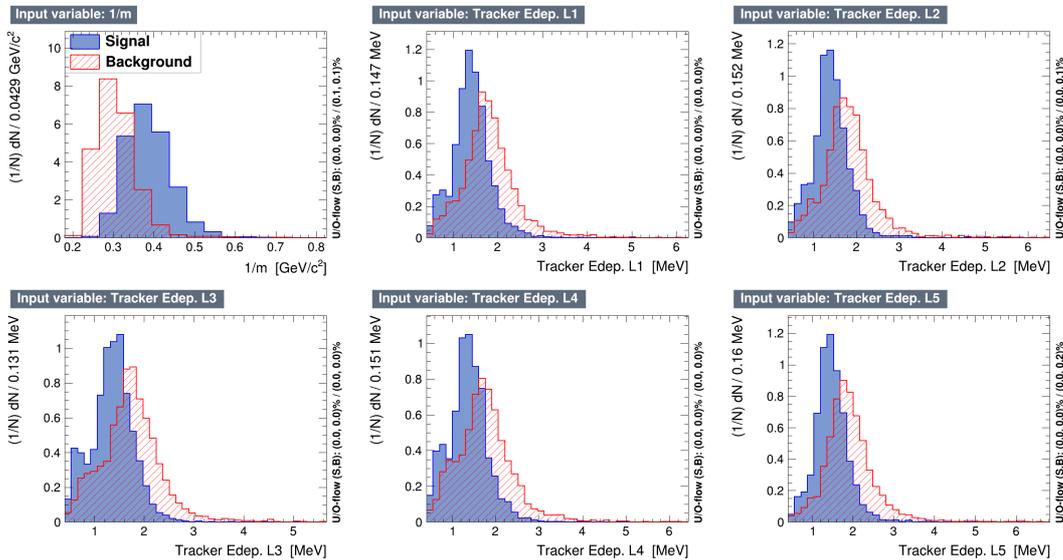


Fig. 3 Example of distribution of training variables in a fixed rigidity between 0.8~1.0 GV/c for helium-4(blue) and helium-3(red). Upper left-most is mass distribution, and the others are energy deposition on each layers of tracker and TOF.

In case of we trying distinguish isotope, i.e., when the particle species we want to discriminate has same electric charge, but with different mass, their energy deposition as a function of rigidity will show slightly different behavior in the region of rigidity less than 10

GV. But, since the both signal and background has same electric charge, the energy distribution is not clearly separable. We presented the comparison of training parameters of signal sample and background sample at Fig. 3.

3. Result

After the training, we evaluated independent event samples to test the performance of BDT. In the next two subsection, we will discuss about two different cases, one is signal and background events have different charge z and the other is for the same charge.

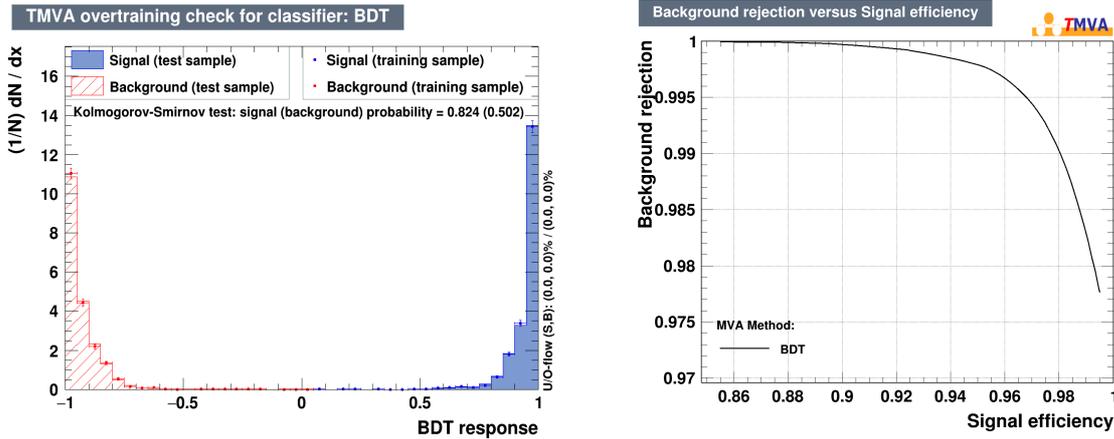


Fig. 2 (a) (left) BDT response after the evaluation. (b) (right) ROC curve of evaluation.

First, we considered different kind of nuclei for signal and background samples. In this case the energy deposition shows very distinctive distribution and it contributes to improving the separation power between signal and background. Fig 4-a shows the result of test and 4-b shows the receiver operating characteristics(ROC) curve and it presenting better background rejection efficiency in a same signal selection efficiency in comparison to the case of cut based method using average energy deposition of tracker and TOF.

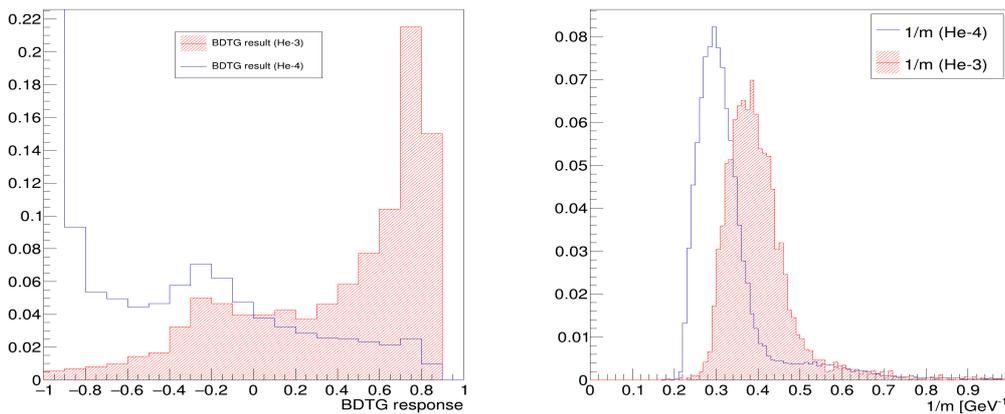


Fig. 3 (a) (left) Evaluation result against helium-3 and helium-4 test sample for isotope separation. This case we chose BDTG for better performance. (b) (right) Mass distribution of helium-3 and helium-4 for comparison.

The second case is the signal and background samples are composed nuclear isotopes of each other. (i.e., helium-3 for signal, helium-4 for background) Fig. 5-a shows the result of test and Fig. 5-b for comparison to the case of using mass information to distinguish two samples.

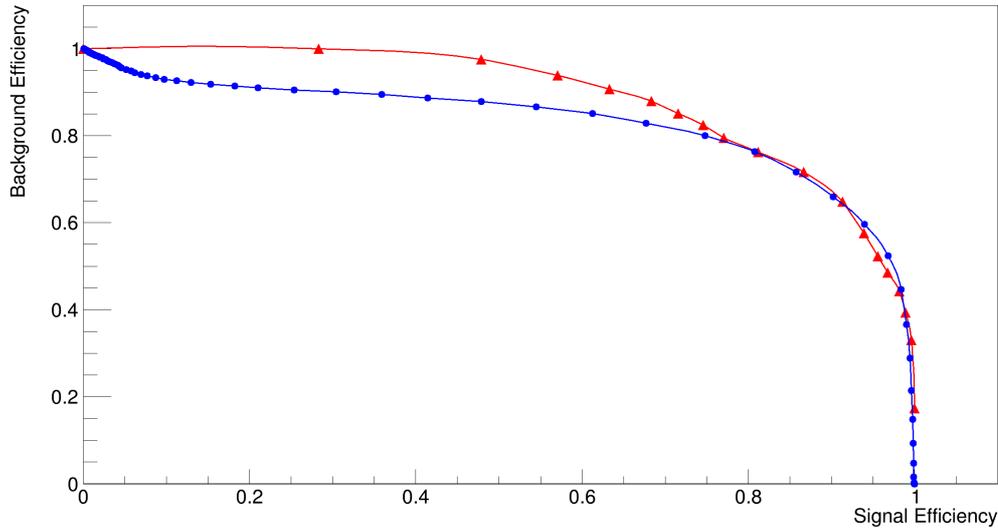


Fig. 4 Receiver Operating Characteristic(ROC) curve of BDTG response and mass distribution. Signal efficiency greater than 0.8 region, both estimators shows no difference but below the signal efficiency 0.8, BDTG estimator gives better background rejection efficiency.

4. Conclusion

We reported our study of multivariate approach against various kinds of light nuclei isotopes of cosmic-rays using simple Monte Carlo simulation that mimics contemporary standard magnetic spectrometers such as AMS-02 and PAMELA. When the incident particle for signal and background has different electric charge, multivariate approach promisingly showed better separation of signals out of backgrounds, especially for this study, we achieved 99% of background rejection efficiency when we required 98% signal efficiency. In case of differentiating nuclear isotopes, multivariate approach showed us similar performance compare to the classical method using reconstructed mass. However we could achieve better background rejection efficiency if we require signal efficiency less than 80%.

Since this study is completely relying on Monte Carlo simulation, we have to study the systematic effect and robustness of the event selection together with evaluation of the BDT estimator against real cosmic-ray data.

References

- [1] S. Haino, et. al., *Performance of the AMS-02 silicon tracker in ISS mission*, Nuclear Instruments and Methods A 699 (2013) 221-224
- [2] O. Adriani, et. al., *The magnetic spectrometer of the PAMELA satellite experiment*, Nuclear Instruments and Methods A 511 (2003) 72-75
- [3] V. Bindi, et. al., *The time of flight detector of the AMS-02 experiment on the international space station*, Nuclear Instruments and Methods A 718 (2013) 478-480
- [4] W. Menn, et. al., *The ToF System of the PAMELA Experiment: In-orbit Performance*, in proceedings of 30th International Cosmic Ray Conference.

- [5] O. Adriani, et. al., *The silicon tracker of the PAMELA experiment*, Nuclear Instruments and Methods A 409 (1998) 447-450
- [6] S. Agostinelli et al., Geant4 - A Simulation Toolkit, Nuclear Instruments and Methods A 506 (2003) 250-303
- [7] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA - Toolkit for Multivariate Data Analysis*, PoS ACAT 040 (2007), arXiv:physics/0703039
- [8] H. Yang, B. Roe, J. Zhu, *Studies of Boosted Decision Trees for MiniBooNE Particle Identification*, Nuclear Instruments Methods A 555 (2005) 370-385
- [9] G. Galluci, et. al., *Performance of the AMS-02 Electromagnetic Calorimeter in Space*, in proceedings of 16th International Conference on Calorimetry in High Energy Physics (CALOR 2014) 587 (2015) 012028.
- [10] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen, *Classification and Regression Trees*, Wadsworth, Stanford, 1984.
- [11] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016) and 2017 update.