Application of an MT method to CALET e/p separation

June 25, 2015

Kenji Yoshida
Event classification and background rejection for cosmic-ray observations

• There are various data mining techniques:
  – Artificial Neural Network, Boosted Decision Tree, etc.
• A Mahalanobis Taguchi (MT) method is an alternative to these data mining techniques.
Mahalanobis Taguchi (MT) method

- Constructing one standard multidimensional unit space from characteristic variables of normal data (e.g., electron data)
- Applying test data to the unit space
- Calculating each Mahalanobis distance of test events on the unit space

- Events with Mahalanobis distances close to 1 are similar to the normal data
- Events with the distances much larger than 1 are dissimilar to the normal data
How to calculate the Mahalanobis Distance

**Mahalanobis distance**

\[ D^2 = \frac{1}{k} Z_i^T C^{-1} Z_i \]

where \( Z_i \) = standardized vector obtained by values of \( X_i (i = 1, \ldots, k) \)

\[ Z_i = \frac{(X_i - m_i)}{s_i} \]

\( X_i \) = value of \( i \)-th characteristic

\( m_i \) = mean of \( i \)-th characteristic

\( s_i \) = standard deviation of \( i \)-th characteristic

\( k \) = the number of characteristics / variables

\( T \) = transpose of the vector

\( C^{-1} \) = inverse of the correlation matrix
Procedure

• In the application of an MT method to CALET e/p separation, electron data are used to construct the unit space as normal data.

• The normal data of electrons correspond to “training data” in machine learning algorithm.
  – We can “train” without proton data.

• In addition, by using proton data, it is possible to select significant characteristic variables.
Selection of characteristic variables

• Using an orthogonal array and SN ratio

   SN ratio: \( \eta(dB) = -10 \log \sigma^2 \)

   \[ \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{D_i^2} \]

• Evaluating the contribution of each attribute and selecting the characteristic variables
  
  — Proton events are used for the selection.
Simulation Data and Preselection

- CALET CAD model Rev.15 Epics simulation data
- Electron 1TeV, Proton E^{-2.7} (>1TeV) (dpmjet3)
- Preselection:
  - High Energy Shower Trigger
  - Particle incident direction inside acceptance (Geometric conditions A-D)
  - TASC Energy Deposit inside energy bin (903-975GeV) to be 95% efficiency
Characteristic variables in this study

- Lateral Spreads of CHD, IMC, and TASC
- Energy Deposit of each CHD, IMC layer
- E.D. Fraction of each TASC layer
- Lateral Spread of each TASC layer
- TASC_rms^2*E.D.F. of each TASC layer
- IMC longitudinal profile fitting => p0, p1, red-\chi^2
- TASC longitudinal shower profile fitting => t_{\text{max}}, b, red-\chi^2

\[
\frac{dE}{dt} = p_0 t^2 + p_1
\]

\[
\frac{dE}{dt} = \frac{E_0 b}{\Gamma(b t_{\text{max}} + 1)} (b t)^{b t_{\text{max}}} e^{-b t}
\]
Geometric Conditions: A-D

Applying the MT method for each geometric condition
e/p separation: Geometric Condition A

- For constructing the Unit Space:
  2386 electron events
- For selecting variables:
  336 proton events

- Test sample:
  3970 electron events
  1801 proton events

- In the case of the threshold $D = 1.10$
  Survived electron events: 3419
  Survived proton events: 0

Selected variables

<table>
<thead>
<tr>
<th>SN-&lt;SN&gt;</th>
<th>TASC12_E.D.Fraction</th>
<th>0.693</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TASC12_rms^2*E.D.F.</td>
<td>0.663</td>
</tr>
<tr>
<td></td>
<td>TASC11_rm2^2*E.D.F.</td>
<td>0.304</td>
</tr>
<tr>
<td></td>
<td>TASC11_E.D.Fraction</td>
<td>0.251</td>
</tr>
<tr>
<td></td>
<td>TASC9_E.D.Fraction</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>TASC1_rms^2</td>
<td>0.170</td>
</tr>
</tbody>
</table>

... (18 variables in total)
e/p separation: Geometric Condition B

- For constructing the Unit Space:
  - 1681 electron events

- For selecting variables:
  - 313 proton events

- Test sample:
  - 1763 electron events
  - 918 proton events

- In the case of the threshold $D = 1.23$
  - Survived electron events: 1607
  - Survived proton events: 3

Selected variables |
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TASC12_E.D.Fraction</td>
</tr>
<tr>
<td>TASC12_rms^2*E.D.F.</td>
</tr>
<tr>
<td>TASC11_rm2^2*E.D.F.</td>
</tr>
<tr>
<td>TASC11_E.D.Fraction</td>
</tr>
<tr>
<td>IMC_fit_p0</td>
</tr>
<tr>
<td>CHD_E.D.</td>
</tr>
</tbody>
</table>

... (11 variables in total)
**e/p separation: Geometric Condition C**

- For constructing the Unit Space:
  - 1302 electron events
- For selecting variables:
  - 299 proton events

- Test sample:
  - 1460 electron events
  - 614 proton events

- In the case of the threshold \( D = 1.21 \)
  - Survived electron events: 1341
  - Survived proton events: 0

### Selected variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>SN-&lt;SN&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>TASC12_E.D.Fraction</td>
<td>0.891</td>
</tr>
<tr>
<td>TASC12_rms^2*E.D.F.</td>
<td>0.717</td>
</tr>
<tr>
<td>TASC11_rms^2*E.D.F.</td>
<td>0.408</td>
</tr>
<tr>
<td>TASC11_E.D.Fraction</td>
<td>0.370</td>
</tr>
<tr>
<td>TASC10_rms^2*E.D.F.</td>
<td>0.223</td>
</tr>
<tr>
<td>TASC9_rms^2*E.D.F.</td>
<td>0.197</td>
</tr>
</tbody>
</table>

... (12 variables in total)
e/p separation: Geometric Condition D

- For constructing the Unit Space:
  1875 electron events
- For selecting variables:
  345 proton events

- Test Sample:
  2358 electron events
  1172 proton events

- In the case of the threshold D = 1.20
  Survived electron events: 2079
  Survived proton events: 3

---

### Selected variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>SN-&lt;SN&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>TASC11_rms^2*E.D.F.</td>
<td>0.523</td>
</tr>
<tr>
<td>TASC12_rms^2*E.D.F.</td>
<td>0.473</td>
</tr>
<tr>
<td>TASC12_E.D.Fraction</td>
<td>0.447</td>
</tr>
<tr>
<td>TASC11_E.D.Fraction</td>
<td>0.278</td>
</tr>
<tr>
<td>TASC9_rms^2*E.D.F.</td>
<td>0.226</td>
</tr>
<tr>
<td>TASC10_E.D.Fraction</td>
<td>0.212</td>
</tr>
</tbody>
</table>

... (25 variables in total)
Proton Rejection Power

Proton Rejection Power = Incident protons $1.03 \times 10^6$/Survived protons

Electron Survival Rate = Survived electrons divided by electrons $9.55 \times 10^3$ in TASC energy bin 903-975 GeV

The same threshold distance is used in geometric conditions A-D
Summary

• Application of an MT method to CALET e/p separation:
  – For the training, the number of electron events is relatively small, possibly without proton events.
  – e.g. Proton Rejection Power is $1 \times 10^5$ at the electron survival rate of 90%.

• Refining the merit variables for the MT method application should be in progress.