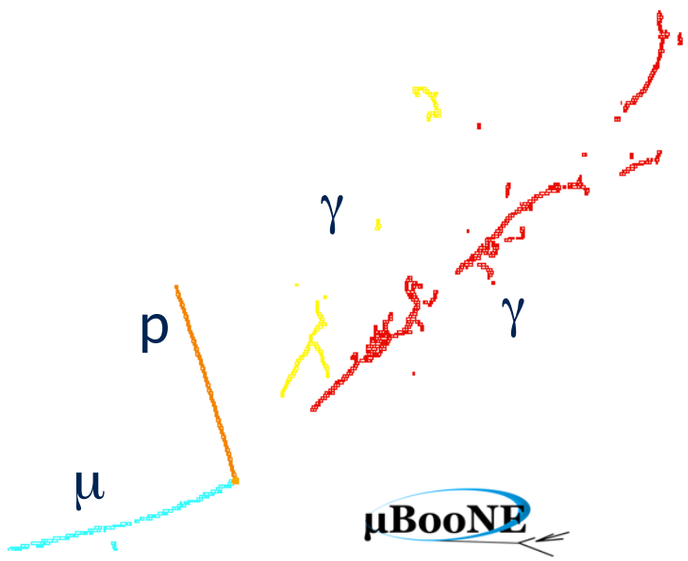


Recent updates on SVM-based track/ shower identification in Pandora

L. Escudero for the Pandora team

LArSoft Coordination Meeting

13th of February, 2018

Outline

This talk outlines the changes proposed for

larpandoracontent v03_10_00
and related for
larpandora, dunetpc
uboonecode, uboonedata

to incorporate the recent updates on the
SVM-based track/shower classification

This is a small update concerning one existing algorithm
and involving only minor changes

Introduction

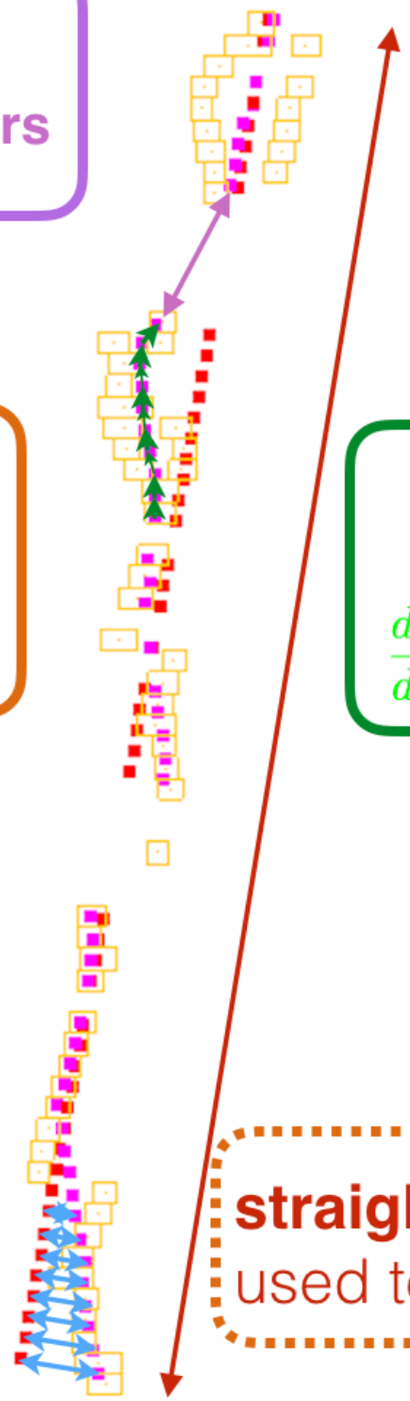
- A new algorithm was added in larpandoracontent v03_07_00 (June 2017) for track/shower classification based on Support Vector Machines (SVMs) for use by analyzers downstream
- The algorithm uses the SVM mechanics shared with the SVM-based vertex selection algorithm, i.e. abstracted by the SVM object and associated helper class
- The classification is performed based on a number of features (physical variables)
- This algorithm runs at the end of the reconstruction chain: it does not affect the pattern recognition, but only the final label assigned to PFOs, whether they are more likely to be track-like or shower-like, driving the creation of a recob::Track or recob::Shower associated
- Initially performing a binary classification, then added a possibility of a probability/score to be computed and used
- Trained models have been provided for MicroBooNE as .xml files located in uboonedata

Introduction

Previous SVM features

max separation between fit layers

RMS of the sliding linear fit
sum in all layers



True shower hits

dTdL width
width of fit dT/dL

$$\frac{dT}{dL}_{width} = \frac{dT}{dL}_{max} - \frac{dT}{dL}_{min}$$

shower fit width
sum (layer by layer) difference between fits to positive and negative edges

$$\sum dT = \sum |T_P - T_N|$$

difference with straight line
mean and σ of

$$dT = |T_s - T_L|$$

short/Large fit window

straight line length
used to normalise

distance to vertex

Updates

What has changed:

- List of features used:
  - moved to a 3dimensional approach
- Classification done at the PFO level, as opposed to at the cluster level before
- Probability/score used in the classification
- New trained models provided (.xml files)

What has not changed:

- Same SVM mechanics used
- The algorithm is run at the end of the reconstruction, not affecting pattern recognition but improving the final label given to PFOs for use of analyzers
- Previous version (2D, cluster) can still be used

Updates

What has changed:

- Algorithm and associated tool updated in larpandoracontent
- Changes accordingly in configuration settings files in uboonecode
- New trained model files added to uboonedata (2 files of 9.3 + 1.1 MB)

Additionally:

- Intensive testing has been carried out, using with results from BDTs
- Cross-validation tests done to successfully discard under and overtraining

Updates

New 3D-related features

- **Topological variables:**
  - **Length from ThreeDSlidingFitResult**
  - **Average of normalised diffWithStraightLineMean in all 2D clusters**
  - **Average normalised max fit gap length all 2D clusters**
  - **Average normalised RMS sliding linear fit in 2D clusters**
- **Vertex distance: 3D distance from PFO vertex to nu vertex**

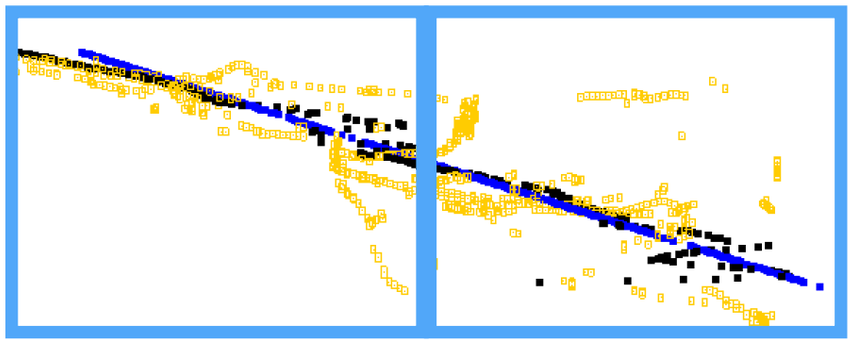
These variables are adapted from the features previously used, either using 3D distances or averaging those 2D variables calculated from all the clusters available for the PFO under study. Topological variables normalised by cluster length for better performance (e.g. a long cluster might present a large total deviation from straight line consequence of adding many small contributions)

Updates

New 3D-related features

- Topological variables:
  - Length from ThreeDSlidingFitResult
  - Average of normalised diffWithStraightLineMean in all 2D clusters
  - Average normalised max fit gap length all 2D clusters
  - Average normalised RMS sliding linear fit in 2D clusters
- Vertex distance: 3D distance from PFO vertex to nu vertex
- PCA
  - Ratio secondary eigenvalue / principal eigenvalue
  - Ratio tertiary eigenvalue / primary eigenvalue
  - Opening - closing angle

select 50% hits closer (further) from vertex, run PCA, find opening (closing) angle



New variables added taking advantage of the Eigen dependency in Pandora added since larpandoracontent v03_07_00 (June 2017)

Updates

New 3D-related features

- Topological variables:
  - Length from ThreeDSlidingFitResult
  - Average of normalised diffWithStraightLineMean in all 2D clusters
  - Average normalised max fit gap length all 2D clusters
  - Average normalised RMS sliding linear fit in 2D clusters
- Vertex distance: 3D distance from PFO vertex to nu vertex
- PCA
  - Ratio secondary eigenvalue / principal eigenvalue
  - Ratio tertiary eigenvalue / primary eigenvalue
  - Opening - closing angle
- Charge variables (only from W cluster)
  - Sigma charge normalised over total charge
  - End charge (last 10% hits) normalised over total charge

Additional charge related variables included. Since they are only calculated using the cluster in the W view, a backup trained model needs to be provided for those PFO's without W cluster, using the rest of features only

Release plan

Package	Changes	Plan
larpandoracontent	Updates in algorithm and tool for SVM Pfo characterisation	Feature branch larpandoracontent_v03_10_00 available on redmine
larpandora	Cosmetic update for consistency	Feature branch larpandoracontent_v03_10_00 available on redmine
uboonecode	Updates in Pandora settings .xml files	Feature branch larpandoracontent_v03_10_00 available on redmine
dunetpc	Cosmetic update for consistency	Feature branch larpandoracontent_v03_10_00 available on redmine
uboonedata	Added new .xml files trained models in PandoraData	Feature branch larpandoracontent_v03_10_00 available on redmine

Summary of feature branches

Features branches with name
feature/larpandoracontent_v3_10_00
exist for the following packages:
larpandoracontent,
larpandora, dunetpc
uboonecode, uboonedata

The feature branches are ready, already available in redmine and have been tested for MicroBooNE and DUNE in LArSoft v06_67_01