Institute for High Energy Physics

Some Thoughts about Event Reconstruction with ML in HK

IFAE Neutrino Group

Outline

Institute for High Energy Physics

- Disclaimer
- Motivation
- SK Reconstruction Chain
- Some ideas for ML subtasks (not complete)
- What could be a ML reconstruction chain?
- How to compare performances?
- How to handle uncertainties?
- Conclusions

Disclaimer

Institute for High Energy Physics

This talk is meant to trigger discussions about a possible reconstruction chain using Machine Learning and possibly classical techniques!

Only some thoughts and ideas are presented, there is no claim that all our considerations are correct and also no intention to present a full concept of reconstruction chain!

Motivation

Institute for High Energy Physics

- IFAE started to work on ML in HK few months ago
- 2 starting projects:
  - Iñaki: e/γ separation using ResNet18 adding hit time information => Master thesis: 7% improvement ACU, 16% in electron efficiency
  - Annalisa: Single vertex/multi vertex classification using ResNet18 with q,t information
- Triggered a lot of internal discussions e.g.:
  - e/γ separation with vertex time instead of hit time more efficient?
  - SV/MV classification seems very complicated with hit info but should be trivial if rings/tracks are reconstructed before
  - What we do with SV/MV classification afterwards? Better to know number of vertices? Or output should be rough 4D position of vertices?
- Perhaps we complicate our lives by insisting on using hit information for reconstruction?

FitQun and SK Reconstruction

Institute for High Energy Physics

- Started to look into FitQun and SK reconstruction
- Some lectures taken from this:
  - Low energy events: Bonsai, high energy events: FitQun
  - Hits can be: direct light, indirect light, dark counts
  - Pre-vertex fit using all raw hits to find rough vertex position in 4D
  - Corrections for e.g. PMT angular acceptance applied
  - Peak finder and hit time clustering which returns number of peaks and corresponding times
  - Ring fitter with 7 variables: vertex position ***x***, time ***t***, zenith angle and azimuth of the direction ***theta/phi*** and momentum ***p*** for different particle hypothesis
  - PID based on likelihood value

Possible ML Reco Chain Subtasks?

Institute for High Energy Physics

- PreVertex Fit: Could be used from FitQun
- Hit Classifier:
  - Using eg GNN to classify in direct ring hits, indirect and noise/hidden rings?
  - Seems Matt Lowe had same idea for Bonsai
  - Input: raw hits, with corrected vertex time?
  - Output: classification for each hit allowing to separate them in groups for analysis
- Ring finder:
  - CNN? GNN?
  - One for high energy (multi-)rings and one for low energetic hidden rings?
  - How to deal overlapping rings?
  - Input: All raw hits 2D image? Only direct ring hits? Only noise/hidden rings?
  - Output: List of hits for each ring

Possible ML Reco Chain Subtasks?

Institute for High Energy Physics

- Particle ID:
  - Already successful tests with CNN it seems
  - Worth to try with GNN perhaps?
  - Input: Ring hits
  - Output: classification probability
- Ring Analyser:
  - Classical approach from FitQun?
  - ML for all 7 parameters at once or dividing in subtasks?
  - E.g. Particle-typical CNN for better vertex position?
  - Input: ring hits, PID info?
  - Output: the 7 ring parameters?

Possible ML Reco Chain Subtasks?

Institute for High Energy Physics

- Multi Vertex finder:
  - CNN? Other architecture?
  - No distinguishing between decay chain vertices and pile up at this stage
  - Input: Raw hits? Hits with corrected vertex time? Ring vertices?
  - Output: Number of vertices in event? Rough 4D position of every vertex?
- Pile-up/Decay chain classifier:
  - BDT? Other architecture?
  - Input: Rough 4D position of each vertex? Also PID information? Energy estimation? E.g. two muons from different vertices increase probability that pile up
  - Output: List of vertices belonging to one neutrino interaction?

Possible ML Reco Chain Subtasks?

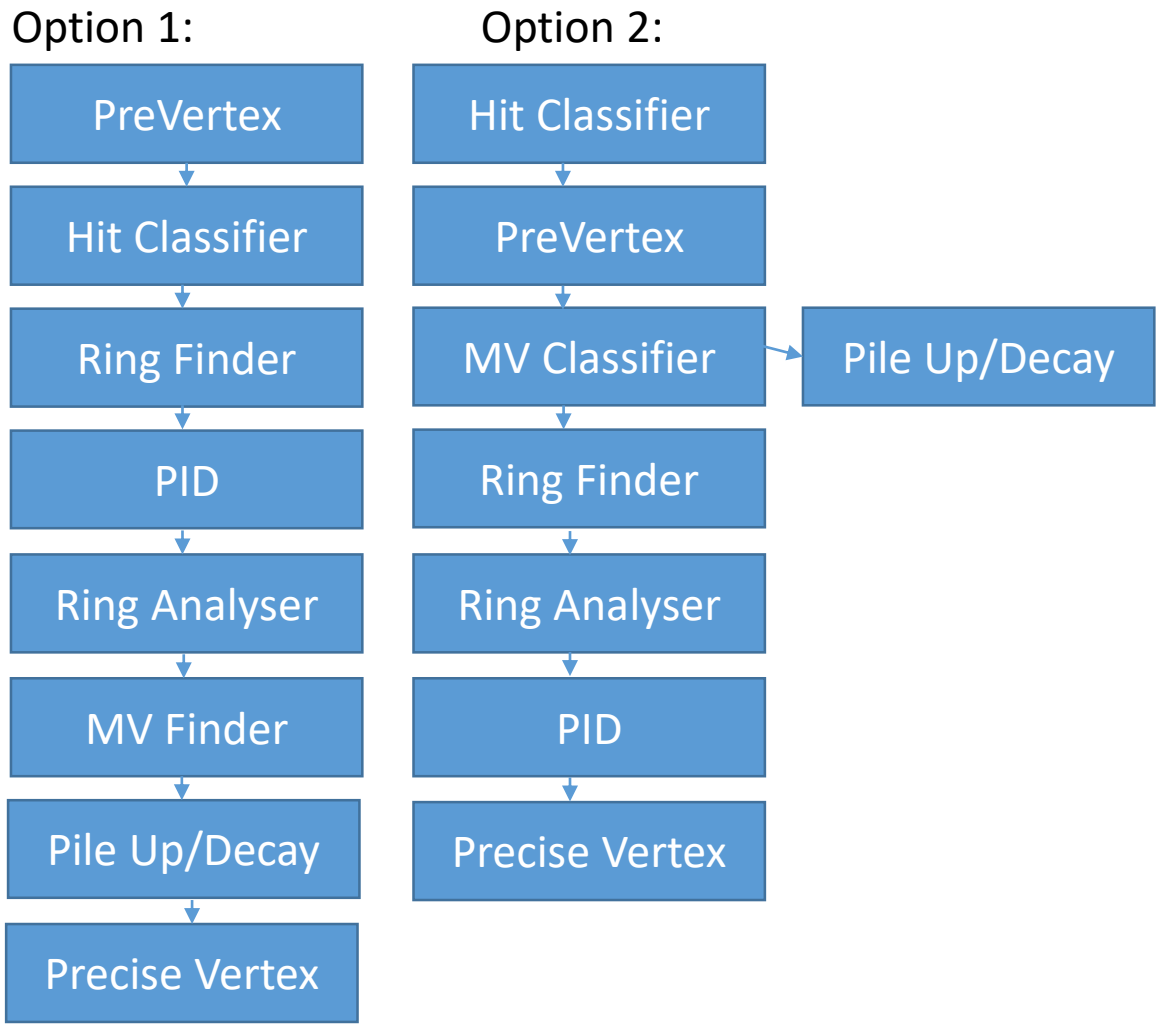
Institute for High Energy Physics

- Precise Vertex Position:
  - Calculate precise vertex position from all hits or ring vertices of belonging to one subevent
  - Classical fit method? ML?
  - Input: calibrated, time corrected hits? Ring vertex positions?
  - Output: 4D position of subevent vertex
- Other Subtasks:
  - Total energy estimation
  - Hit correction e.g. for light scattering
  - ….

ML Reconstruction Chain Options

Institute for High Energy Physics

- As implied in previous slides, it would be good to define roughly position of each subtask in reco chain
- Not fully fixed but to have ideas about possible ordering of subtasks
- Would allow to provide guidance for which input could be expected and which output info should be provided
- Would also allow that anybody could pick a subtask from the list and use the ML architecture they consider useful



Performance Comparison

- Groups might try different approaches for the same subtask
- Would be good to have some criteria to compare performances
- Might not be trivial:
  - Iñaki's study: AUC 7% better but e ring efficiency improved by 16% with 3 million events
  - Specific number of samples for training, verification and testing?
  - Time it takes to analyse/classify one event?
  - Same DAC file settings? DAC threshold seems to have big impact on MC data.

Iñaki:

Model	Loss	Accuracy	F_1 score	AUC
Proposed	0.5856	0.6864	0.6795	0.7550
Original	0.6271	0.6427	0.6203	0.7007

Table 2: Loss function evaluated in the test set along with some performance metrics for the original configuration and the proposed model

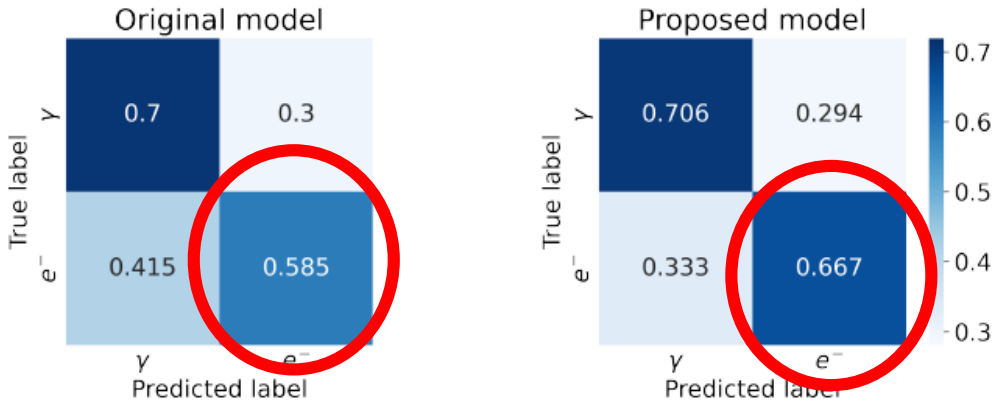


Figure 16: Row normalized confusion matrix for both original and proposed models

Uncertainties

- There will be obvious uncertainties to be dealt with
- Would be good to define what should be considered before working on the subtasks?
- Different types of uncertainties to be considered:
  - For example uncertainty of PE in each PMT => same testing sample with smearing signal in each PMT using same ML?
  - Uncertainty due to ML architecture? Iñaki tried MC dropout for this uncertainty. Other options?

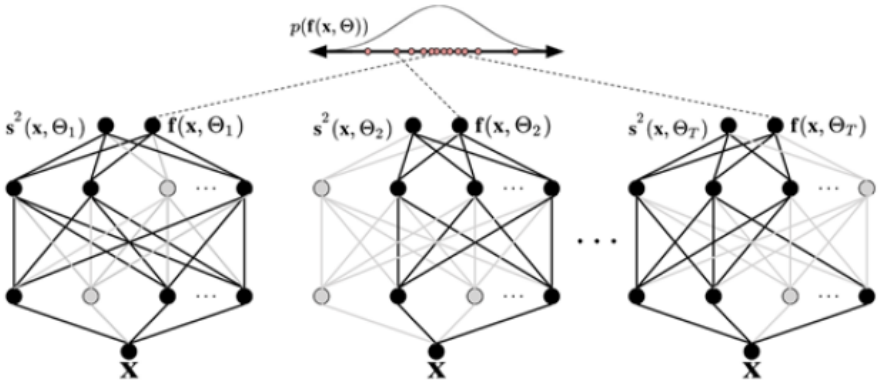


Figure 11: MCD concept [2]

Conclusions

Institute for High Energy Physics

- The two starting IFAE starting projects are very useful
- Thanks for all the input we received
- Remaining doubts: What will be the input info for each task? What should be the output? How to bring all this to a reconstruction chain to do physics?
- Conceptual design of reconstruction chain would be useful
- Also defining performance comparison parameters and needs for uncertainty handling would be good!
- Ideal would be to have a technical note summarizing the strategy towards a reconstruction chain including ML steps
- Might also help to minimize the needed computing resources