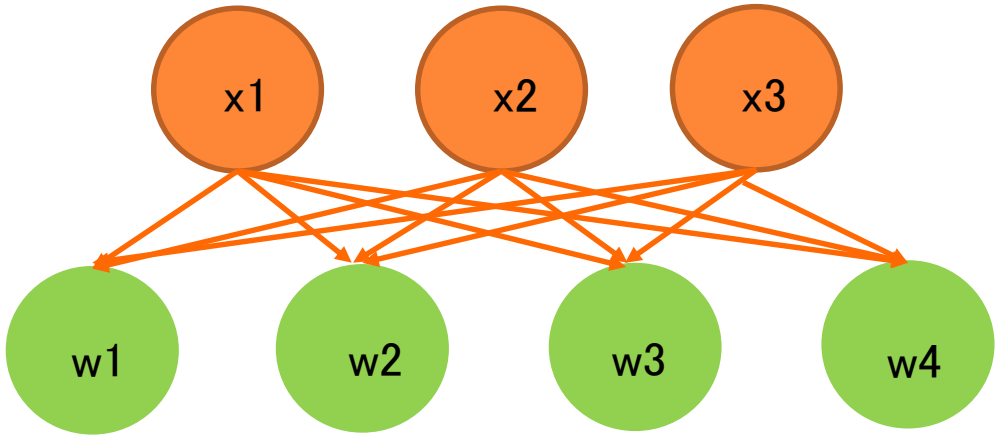
Study For better Jet Clustering using Self-Organized Mapping

Masakazu Kurata

11/27/2015

Introduction

- SOM is one of an "unsupervised" neural network which consists of 2 layers(input and output layers)
  - Input nodes are input data
  - Output nodes have weight vectors
- Looking for "winner" nodes
  - choose output node whose weight vector is similar with input data
  - This node is called as best matching unit(BMU)
- Adjusting the weight vector towards the input vector
  - Both BMU and the nodes which are close to BMU
  - Update formula: $w_j(t+1) = w_j(t) + \alpha_t(x - w_j(t)),$
  - α_t: neighborhood function(distance between BMU and other nodes)



Input Layer

Output Layer

Arrangement for jet clustering

- Output vectors
  - Weighted sum of input vectors $P_i = \sum w_{ij} \cdot p_j , 0 \leq w_{ij} \leq 1$
  - Sum of weight is 1 $\sum_i w_{ij} = 1$
  - So, if num. of output nodes is same as num. of clustered jet, Pi is a (reference) jet 4-momentum
- Looking for "winner" node
  - The output node which has smallest distance measure

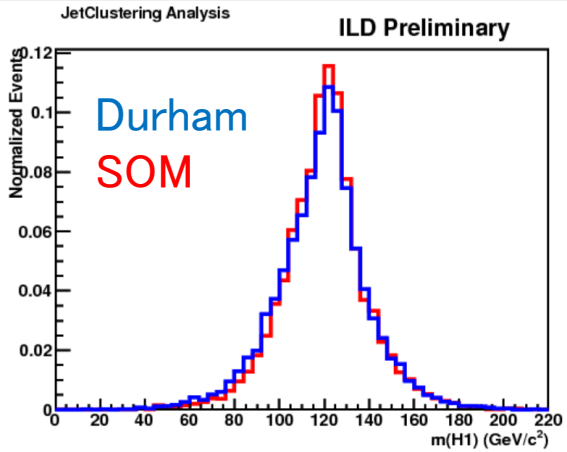
    e.g.) smallest Durham
- Distance between output nodes
  - Using distance measure like Durham y-value
- Update strategy
  - Change wij
  - Increase the wij of BMU and close nodes from BMU, and impose weight sum constraint
    - $w_{ij}(t+1) = w_{ij}(t) + \alpha_t(1.0 - w_{ij}(t))$

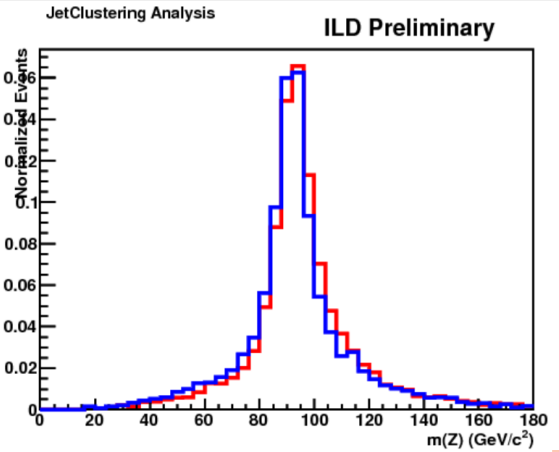
Procedure of SOM for jet clustering

1. Track is clustered until certain y−value using Durham
   - Each mini−jet is a input data for SOM
2. Default output vectors are result of Durham jet clustering
3. Choose BMU for a mini−jet p_j
   - Using distance measure of Jade with output vectors
4. Update weights of BMU and neighborhood output nodes
   - $w_{ij}(t+1) = w_{ij}(t) + \alpha_t(1.0 - w_{ij}(t))$
   - Using distance measure of Jade between BMU and other output nodes
   - αt shrinks gradually with epoch
5. 3. and 4. are performed for all the input data(mini−jets)
6. 5. is performed many times(called as “epoch”)
7. After the training, each mini−jet is assigned to the output node which has smallest Jade distance measure
   - Node is regarded as a ‘jet’

Preliminary results

- Using qqhh→qq(bb)(bb): 6 jet clustering(6 output nodes)
  - Use same event as original Durham clustering
  - Jet matching with MC truth is performed
  - Mass distribution:

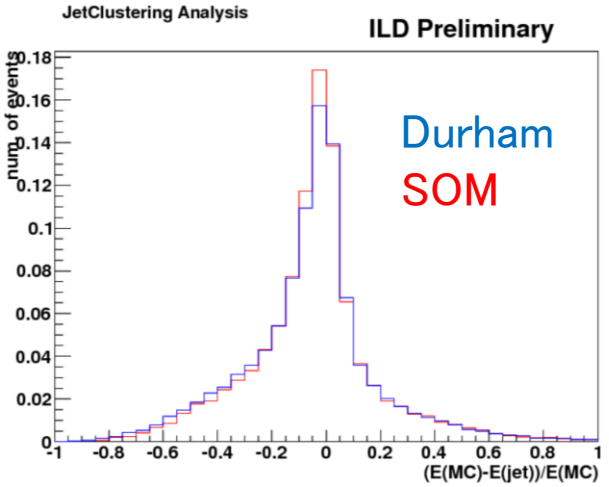




  - Num. of MC matched events:

Method	Durham	SOM
Num. of events	6178	6253

  - Jet Energy resolution of bjets



Further trial

- SOM seems to fall into local minimum when default output nodes are originally mis-clustering result
- If default output vectors describe relatively correct pattern, how is the jet clustering result?

- Trial:
  - Perform 8 jet clustering
  - 6 jet clustering is performed using mass constraint

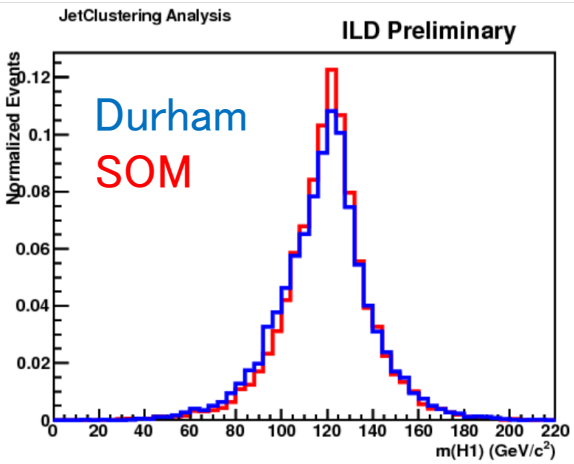
$$\chi^2 = \frac{(m_1 - m_H)^2}{\sigma^2} + \frac{(m_2 - m_H)^2}{\sigma^2} + \frac{(m_3 - m_Z)^2}{\sigma^2}$$

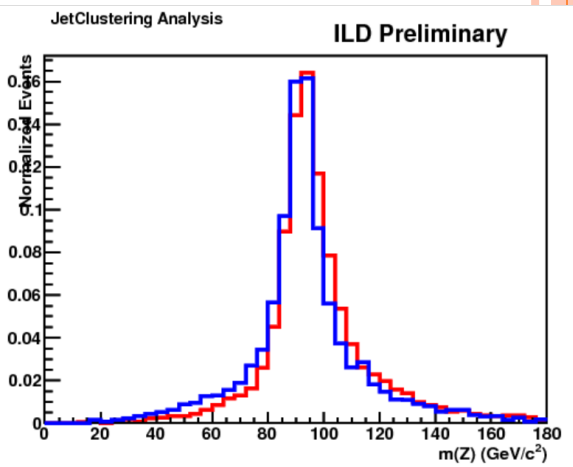
  - Using these vectors as output vectors, SOM is performed in same way

  - 8→6 is based on Junping's study result
  - Not yet tried 10→6, 12→6 etc.

Preliminary results

- Using qqhh→qq(bb)(bb): 6 jet clustering(6 output nodes)
  - Use same event as original Durham clustering
  - Jet matching with MC truth is performed
  - Mass distribution:

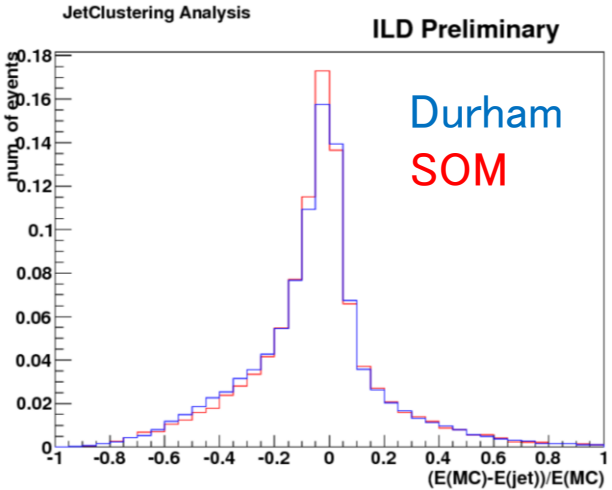




  - Num. of MC matched events:

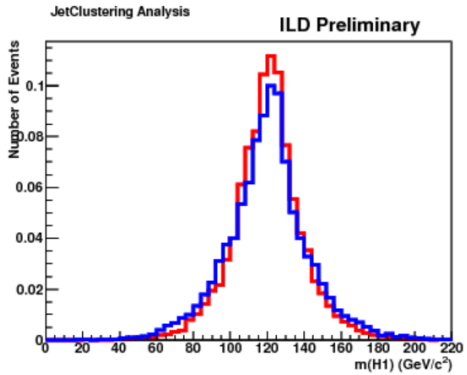
Method	Durham	SOM
Num. of events	6178	6100

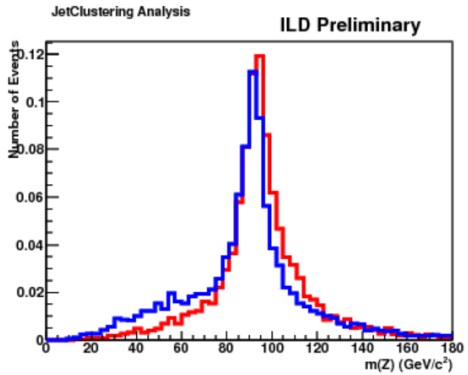
  - Jet Energy resolution of bjets

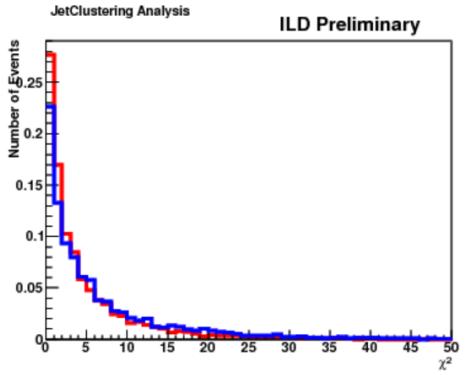


Realistic case

- Trying realistic case for analysis qqHH→qq(bb)(bb)
  - 6 jet clustering
  - Btag>0.35 for 4 jets, no jet energy constraint
  - Mass constraint of Higgs and Z is imposed(χ 2)
  - No MCTruth information is used







Durham
SOM

- Num. of events
  - b-tagging becomes better thanks to clustering?

Method	Durham	SOM
Num. of events	5743	6075

Problems and prospects

- Problems
  - SOM method doesn't reflect physics perfectly…
    - I don't know why only Jade can obtain such results…
  - I don't know the arrangement for jet clustering is good…
  - In the case of backgrounds, especially ZZH
  - In the case of other processes
    - Can obtain similar result?
  - Physics process specific
- Prospects if OK:
  - Improvement of the choice of default output vectors
    - Using Matrix element? Only mass is not enough?
    - 12→6 for example
  - Optimization of several parameters
    - Neighborhood function
    - Learning rate
    - Mass resolution(σ) for mass constraint
    - Etc.
- Hope for jet clustering improvement??