Friendly guide for performing electron neutrino analysis with WatChMaL framework.

First of all, let's describe the pipeline that we will follow.

WCSim.root → *WCSim.npz* → *WCSim.h5*; *split_WCSim.npz* → (model evaluation) → *softmax_WCSim.npy* → *softmax_WCSim.root*.

So, we have to convert initial WCSim files to h5 format, prepare split_files for the model evaluation, run model, achive softmax files and convert them back to root format. All steps have corresponding runner scripts.

- Step1: *WCSim.root* → *WCSim.npz*
  python convert_wcsim_to_npz.py
  In first step we have to prepare npz files ("npzfiles") for initial root files from WCSim simulation and initialize the structure of data directory and copy fiTQun files in the neighbour directory "fitqunfiles".

- Step2: *WCSim.npz* → *WCSim.h5*
  python convert_npz_to_h5.py
  In this step we create corresponding directories and convert "npz-data" files to "h5data".

- Step3: prepare *split_WCSim.npz*
  module load scipy-stack/2020b
  python make_split_idxs_path.py
  In this step we prepare data.split_path files ("splitfiles") for further evaluation and files.txt with h5 and fitqun listed in them.

- Step4: run model evaluation
  python eval_files.py
  #TODO In this step we should one-by-one evalueate model on h5 files data listed in txt format. Output is gathered in rundir.

- Step5: *softmax_WCSim.npy* → *softmax_WCSim.root*
  python softmax_to_root.py
  This step was done locally. We have to convert softmax files to root format and list them. Yet to list cedar locations of softmax.root files in "softmax.txt".

For running Charlie's analysis you need to have corresponding fitqun and softmax files, where softmax file is an output of evaluation with WatChMaL model on wcsim data. Then you need to list files locations row-by-row in *fitqun_files.txt* and *softmax_files.txt*.

Usage of this files together required changes in **RecoInterface**. Also, Charlie's script *analysis_1ringe.cc*, was modified for next purposes:

- new_first_cut_analysis_1ringe_pmu,
  new_first_cut_analysis_1ringe_pe,
  new_first_cut_analysis_1ringe_pgamma
  for each class probability histos.
- log_first_cut_analysis_1ringe_pmu
  for log scale 1D probability histo.
- analysis_energy and analysis_momentum
  for energy and momentum 1D histos.
- logmomentum_pmu
  for reco momentum with log scale probability mu 2D histo.

For running the code you have to perform next steps:

module load StdEnv/2016
source Source_At_Start_nuPRISM.sh
source $DataTools/cedar_scripts/sourceme.sh

Then for single job run *run_single_job.sh*. Here you can change the scripts you need, e.g.:

./bin/logmomentum_pmu $1 $2 ./outputs/histo_$3.root.
Here arguments correspond to single fitqun_files_N.txt, softmax_files_N.txt and N respectively.

For multiple jobs in parallel just run *run_all_jobs*. It will produce multiple histos in root files, so you need split your softmax and fitqun files (as it was done in np_splited directory).

For converting softmax files to root format run *softmax_to_root.py*. Modify it for proper files locations which would be listed. If you need to split list of all files, use *files_split.py*.

Jupyter notebook contains plots for this analysis.