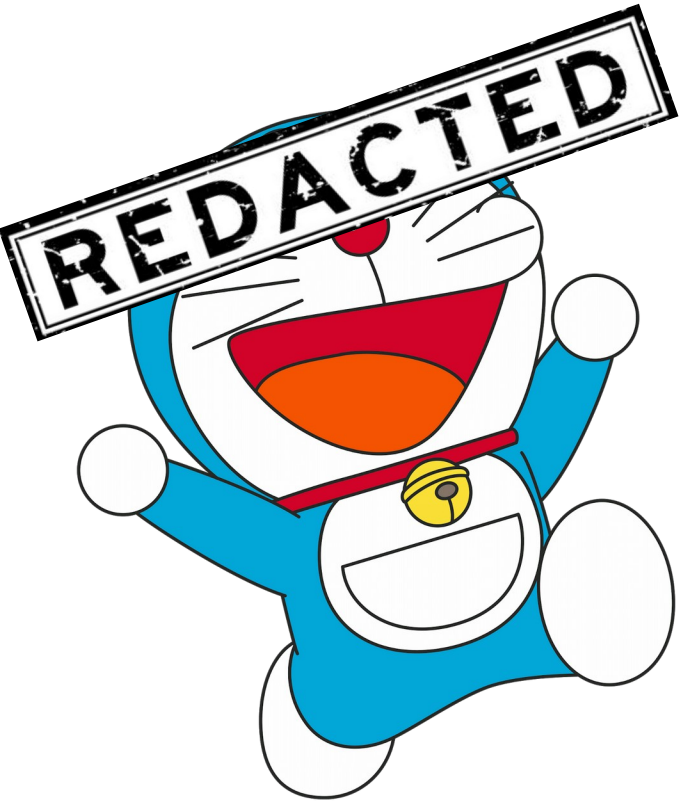
D******N T2 Challenge Planning



Motivation

- 0.1–10 GeV neutrino interaction physics poorly known
- Experiments rely on event generators predictions: GENIE, NuWro, NEUT, GiBUU, Achilles, ...
- Event generation is cheap, but detector simulation is expensive
- Typically can only fully simulate 1–2 configurations
- Risk of inadequately assessing interaction uncertainties for Hyper-K/DUNE

DUNE FD CCINC ν_μ-^{40}Ar

Half a solution

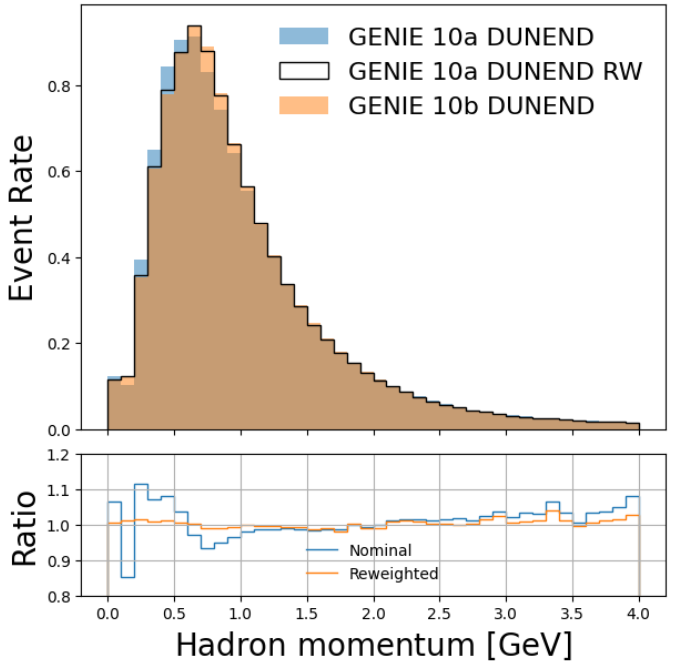
- Try to ***reweight*** between generators
- Train a model to classify between A and B (true label p, prediction q)

$$L_i = p_i \log q_i + (1 - p_i) \log(1 - q_i)$$

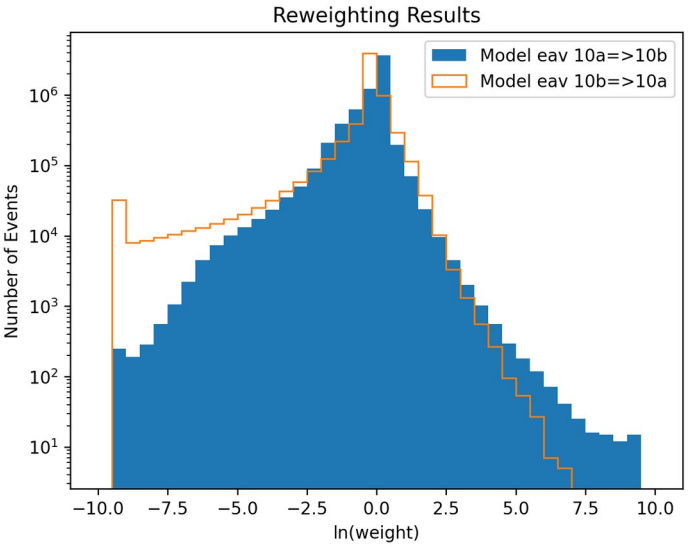
- Use prediction as weights for reweighting

$$\mathcal{L} = p_{\mathrm{A}}(x_i)/p_{\mathrm{B}}(x_i) \approx q_i/(1 - q_i)$$

$$w(\vec{x}) = \frac{G_{\mathrm{B}}(\vec{x})}{G_{\mathrm{A}}(\vec{x})}$$



- This works pretty well!
- But, weights show phase space issue
- Large/small weights also reduces effective stats of reweighted distribution



Plots from Andrew Cudd and Roger Huang

What's the full solution?

- Generate in a wider phase space and then deweight regions which are not there in your target model

$$w(\vec{x}) = \frac{G_{\mathrm{B}}(\vec{x})}{G_{\mathrm{A}}(\vec{x})}$$

This can be 0

This cannot be 0
(for any model of interest)

(This approach currently used in some GENIE tunes to make nuclear binding energy reweightable)

- **Challenges:**
  - Very difficult (impossible) to generalize analytically
  - Deweighting reduces effective MC statistics
- ***General solution:*** *surrogate event generation model*

Task and evaluation

- **Goal:** produce an event sample for a target E_ν distribution which can be reweighted to closely approximate multiple generators
- **Submission:** a sample of (1M) events and 4 vectors of weights, where weights correspond to the 4 generators of interest
  - Event format same as the format we provide events in
  - Weight vector is 1 float/event(/generator)
  - Events correspond to a complex flux, could also have a simpler monoenergetic task *(thoughts?)*
  - We will keep a hold out set of events for evaluation
- **Evaluation:** Energy-distance between submitted *reweighted* events and hold out generator samples (for all 4 generators)

Open datasets: configurations

- Bit of a open question whether there should be multiple configurations for the submissions… for now assume one
- **Generators:** GENIE 10a, GENIE 10b, NuWro v25, NEUT v580
  - More possible, but probably unnecessay
  - All events charged-current only to simplify the problem
  - HDF5 file format ~100 MB/million events
- **Monoenergetic:** simplest case, provided as a test sample
  - ν_μ-^{40}Ar, 0.6, 2.5 and 10 GeV
  - 10 million events each
  - *Could also be a simpler evaluation task?*
- **Experimental flux:** most interesting as challenge datasets
  - DUNE FHC ν_μ-^{40}Ar FD unoscillated and oscillated (0.2–8 GeV)
  - 10–100 million each (we'll see what we need in testing)
  - Could easily provide Hyper-K ν_μ-^{16}O files, but unclear benefit…
- **Flat flux:** essential if a model needs to work on multiple fluxes… but maybe unnecessary for this challenge?

Data format

Event level quantities

- Mode is a bit questionable as generator definitions vary… keep for now
- CC is superfluous as the planned release is CC-only, but prefer to keep...

Dataset	Type	Description
`mode`	int32	Interaction mode
`pdg_nu`	int32	Neutrino PDG code
`pdg_tgt`	int32	Target PDG code
`Enu`	float32	True neutrino energy
`W`	float32	Hadronic invariant mass
`Q2`	float32	Momentum transfer squared
`q0`	float32	Energy transfer
`q3`	float32	3-momentum transfer
`cc`	bool	Charged-current flag
`nfsp`	int32	Number of final-state particles

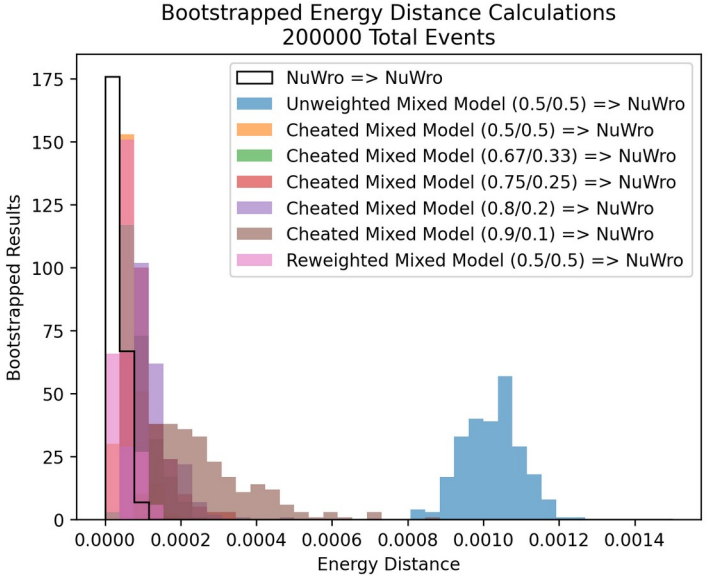
+Full particle stack

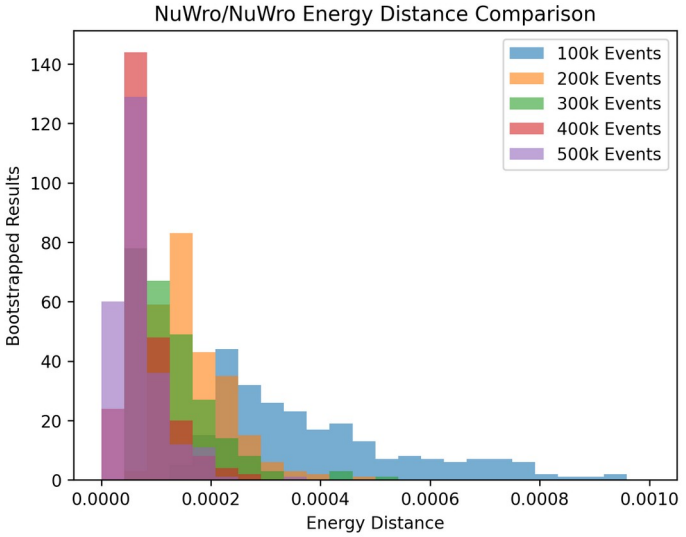
Dataset	Type	Description
`offsets`	int64	Indices into particle arrays; event `i` spans `offsets[i]` to `offsets[i+1]`
`px`	float32	x-momentum
`py`	float32	y-momentum
`pz`	float32	z-momentum
`E`	float32	Energy
`pdg`	int32	PDG code

Currently process with: Generator → NUISANCE → lightweight format, might add it to NUISANCE to reduce steps

Evaluation metric

- Plan is to use energy-distance metric using a reduced set of derived variables: E_ν, E_{avail}, $N\pi^0$, $N\pi^\pm$, Σp_{proton} ($p > 0.1$ GeV)
- These should be hard to cheat, easy to calculate, and representative of what we need for real experiments
- Good for determining how close distributions are, but also interested in how average size of weights (statistical efficiency of reweighted sample)
- *Investigating bootstrap resampling + energy-distance to solve this*





What will we provide?

Plan TBD
Clear plan
In progress
Complete

- Document which describes problem and datasets
- The various datasets described
- +containers and example generation scripts for completeness:
  - https://github.com/wilkinson-nu/t2_doraemon
  - https://hub.docker.com/repository/docker/wilkinsonnu/t2_doraemon
- Evaluation script we will use (without the holdout samples)
- Baseline solution for surrogate model + reweighting
- +scripts for training/testing the above
- Visualization tools etc for the online portal?

Who (so far)?

- **Callum Wilkinson, LBNL:** dataset generation + curation
- **Roger Huang, LBNL:** metric development + evaluation script
- **Vinicius Mikuni, KMI:** surrogate model baseline design + scripts
- **Andrew Cudd, CU (→LBNL):** baseline reweighting solution + scripts

- Other input definitely welcome! The surrogate model design and metric development are currently ill-defined, so let us know if you're interested!