Deep Learning Classifier for Low-Energy Events in IceCube

Maria Prado Rodriguez
for the IceCube Collaboration

July 24th, 2020
mvprado@icecube.wisc.edu
Overview

• DeepCore Subarray
• Neutrino Oscillations
• Event Classification in DeepCore
• 2D CNN DeepCore Classifier
• Future Improvements
### DeepCore Subarray

<table>
<thead>
<tr>
<th></th>
<th>IceCube</th>
<th>DeepCore</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of strings</strong></td>
<td>78 IceCube strings</td>
<td>8 DeepCore strings</td>
</tr>
<tr>
<td><strong>Vertical distance</strong></td>
<td>17 meters</td>
<td>7 meters</td>
</tr>
<tr>
<td><strong>between DOMs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Horizontal distance</strong></td>
<td>110-150 meters</td>
<td>40-90 meters</td>
</tr>
<tr>
<td><strong>between DOMS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total DOMs per string</strong></td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

**Diagram Notes:**
- IceCube string
- DeepCore string
- Ext. DeepCore
- Corridor
- Depth (m)
- Veto cap 10 DOMs 10 m vertical spacing
- Dust layer
- DeepCore 50 HQE DOMs 7 m vertical spacing
- Absorption (m$^{-1}$)
Tighter vertical spacing provides better event resolution

Can probe events with energies as low as 5 GeV

Study “upward-going” neutrinos coming from cosmic rays hitting the Earth’s atmosphere

“Upward-going” means neutrinos that traverse the Earth
Neutrino Oscillations

- Atmospheric muon neutrino oscillations are studied with DeepCore.
- Looks into:
  - Whether $\theta_{23}$ is maximal
  - Unitarity of the PMNS matrix via tau normalization value
- Neutrino energies most relevant for these oscillation studies range from 20 to 35 GeV.
- Events at these low energies are sorted as track-like or cascade-like (no double-cascades).
- Tracks: $\nu_\mu CC$, cascades: every other neutrino event.
- $\nu_\mu \rightarrow \nu_\tau$ oscillations, for example, appear as an excess in the number of cascades observed.
- Therefore, accurate classification is crucial.
Event Classification in DeepCore

- **Current method:** Feed reconstructed quantities into an XGBoost Boosted Decision Tree for learning

- **Goal:** Use deep learning to exploit symmetries in simulated data for classification

- **Problem:** Mapping DeepCore to an input grid proves more challenging than it sounds, given that it does not have good X-Y-Z spatial symmetry

- **One solution:** Map the DeepCore PE charge hits to a 3D grid (t, z, strings) where the string dimension is a channels dimension
2D Event Classifier

- Event images have dimensions:
  \((t, z, \text{strings}) = (40, 49, 15)\)
- Excluded region in ice of increased light-scattering and absorption (Dust Layer)
- Energies > 10 GeV for better resolution
- Events have at least 8 DOM charge hits
- Events start within radius of 100 m from center of DeepCore (orange circle below)

![Diagram of DeepCore and IceCube strings with event images and channels]

Maria V. Prado Rodriguez
2D Event Classifier

• Accounted for the difference in z-axis spacing for DeepCore and IceCube strings
• Padded IceCube strings with zeros in between for more accurate mapping
• Still not a perfect mapping
• Timing windows are 25 nsec apart
• A 2D CNN over time and z (depth)
• Mapping DeepCore subarray hits is more accurate using time and z (depth) than with x and y
• String dimension is a channels dimension: no specific ordering needed
• Cross-channel patterns
CNN with Strings as Channels

- **Cross-channel patterns:** Multi-channel CNN highlights and merges together patterns found in each channel.
- CNN looks for patterns simultaneously in all channels and finds similarities between channels.
- Strings positioned near each other that are hit by an event should exhibit somewhat similar patterns.
- By looking at the learned filter values, the CNN can use these similar patterns as a form of linking together the strings that are close to each other, gaining back some information on the x-y axis position.

---

Cross-channel patterns:

- Multi-channel CNN highlights and merges together patterns found in each channel.
- CNN looks for patterns simultaneously in all channels and finds similarities between channels.
- Strings positioned near each other that are hit by an event should exhibit somewhat similar patterns.
- By looking at the learned filter values, the CNN can use these similar patterns as a form of linking together the strings that are close to each other, gaining back some information on the x-y axis position.
2D DeepCore Images

- Cascade event with a higher energy than that of most DeepCore events
- Highlights patterns better than lower-energy events
- Shows similar patterns in strings hit that are close to each other (magenta stars)
- Sideways “V” is a characteristic pattern of cascades even though some tracks can still exhibit this pattern
2D DeepCore CNN

- CNN output for 1 filter
- Can see the merged sideways “V” in the top-middle section of the output
- The goal is that patterns like these can still be found in events with energies of about 20 to 35 GeV
• Used low-level files (not much pre-processing and no reconstruction)
• Training: ~ 5 million events
  Validation: ~ 584k events
  Test: ~ 233k events
• Tracks: ~ 3.5 million
  Cascades: ~ 2.4 million
• Events are weighted according to the corresponding atmospheric weights
• The performance of the 2D CNN classifier proves to be very similar to that of the BDT classifier
• Perhaps the inaccurate z-axis spacing or the sparse nature of the events limits the CNN from finding more useful symmetries
• Perhaps more training data is needed
2D CNN Classifier:
track cutoff = 0.56

BDT Classifier:
track cutoff = 0.45
2D CNN Classifier: Prediction Distribution

- Prediction distribution for the test sample
- Expected distribution with a peak at 1.0 (track) and no peak at 0 (cascade)
- The network could never be 100% sure of a cascade
- $\nu_\mu CC$ creates a hadronic shower along with a muon
- “Tracks” are technically lollipop-shaped
- A very short track will look just like a cascade
- A very long track will never look like a cascade
- The peak at the center are the events the network could not distinguish at all
Future Improvements

- More accurate z-axis spacing of DeepCore vs IceCube strings
- Perhaps Cumulative charge
- More training data
- Sparse convolutions
- Graph Neural Networks (already in the works)
- IceCube Upgrade: seven more strings with even shorter DOM spacing
- Will help us probe lower energy events (1-10 GeV) by providing more information channels
- Could provide better separation power between tracks and cascades
References

Thank you!
Back-up Slides
The IceCube Neutrino Observatory

- IceCube is a neutrino detector located at the South Pole.
- It consists of a lattice of light sensors embedded in Antarctic ice ~1.5 kilometers below the surface.
- Neutrinos can interact with the atoms in the ice and produce detectable charged particles that leave behind a light trace.
- Our digital optical module light sensors pick up this light, convert it into a current and transfer the digitized current up the cable to the lab.

IceTop
- 81 Stations
- 324 optical sensors

IceCube Array
- 86 strings
- including 8 DeepCore strings
- 5160 optical sensors

DeepCore
- 8 strings-spacing optimized for lower energies
- 480 optical sensors

Vertical spacing of IceCube sensors:
- 17 meters
DOMs are made up mainly of a PMT and the readout electronics. PMTs detect photons through the photoelectric effect, amplifying the photoelectrons knocked off of the photocathode through a dynode chain. By the time the photoelectrons reach the anode, the amplification is big enough to create a readable current.
Cosmic Rays

• High-energy protons and nuclei that collide with the Earth’s atmosphere
• Produce a shower of particles
• Charged pions and kaons from this shower decay into muons and neutrinos
• “Upward-going” direction: Neutrinos will travel through the Earth
• Some will interact with the ice in DeepCore

\[ p^+ + N \rightarrow \pi^+ + X \]
\[ \pi^+ \rightarrow \mu^+ + \nu_\mu \]
\[ \mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e \]

In 2014, we saw about 44,465 neutrino events in DeepCore with an expected ratio of about 5 : 15 : 1 for \( \nu_e : \nu_\mu : \nu_\tau \).
IceCube Events

**Tracks**

\[ \nu_\mu + N \rightarrow \mu^- + X \]

Hadronic shower plus Muon track

*Muons travel long distances before they decay*

**Cascades**

\[ \nu_x + N \rightarrow \nu_x + X \]

Hadronic shower

\[ \nu_e + N \rightarrow e^- + X \]

Hadronic shower plus Electromagnetic shower

*Electrons scatter*

**Double-Bangs**

\[ \nu_\tau + N \rightarrow \tau^- + X \]

Two Hadronic showers

\[ \tau^- \rightarrow \nu_\tau + X' \]

*Taus decay very quickly*
Tau Neutrino Appearance in DeepCore

\[ P_{\nu_\mu \rightarrow \nu_\tau} \approx \cos^4(\theta_{13}) \sin^2(2\theta_{23}) \sin^2 \left( \frac{\Delta m_{32}^2 L}{4E_\nu} \right) \]

\[ L \approx -D_{\text{Earth}} \cos(\theta), \quad \theta > 90^\circ \]

\[ L \approx 0, \text{ otherwise} \]

\[ L = \text{Earth's Diameter} \]

Standard oscillations

E [GeV] \[ L/E (\text{km/GeV}) \]

35 GeV

20 GeV

Probability

\[ \nu_\mu \rightarrow \nu_e \]

\[ \nu_\mu \rightarrow \nu_\mu \]

\[ \nu_\mu \rightarrow \nu_\tau \]
Optimized to look for an excess of tau neutrinos at energy ranges of 5 - 56 GeV for “upward-going” neutrinos

Events are sorted into two categories: “Track-like” and “Cascade-like”

Expected number of tau neutrino events over the uncertainty in the background events shows how likely we are to see a tau neutrino signal per bin
Although most tau neutrino events will have a “Cascade-like” signature, in about 17% of the tau neutrino CC interactions, the tau will decay into a muon and muon neutrino, so this is why some “Track-like” bins appear non-zero.
# Tau Neutrino Appearance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
<th>Analysis $A$</th>
<th>Analysis $B$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neutrino Flux &amp; Cross Section:</strong></td>
<td></td>
<td>Best fit (CC+NC)</td>
<td>Best fit (CC)</td>
</tr>
<tr>
<td>$\nu_e/\nu_\mu$ Ratio</td>
<td>$1.0 \pm 0.05$</td>
<td>1.03</td>
<td>1.03</td>
</tr>
<tr>
<td>$\nu_e$ Up/Hor. Flux Ratio ($\sigma$)</td>
<td>$0.0 \pm 1.0$</td>
<td>$-0.19$</td>
<td>$-0.18$</td>
</tr>
<tr>
<td>$\nu/\bar{\nu}$ Ratio ($\sigma$)</td>
<td>$0.0 \pm 1.0$</td>
<td>$-0.42$</td>
<td>$-0.33$</td>
</tr>
<tr>
<td>$\Delta \gamma_\nu$ (Spectral Index)</td>
<td>$0.0 \pm 0.1$</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Effective Livetime (years)</td>
<td>-</td>
<td>2.21</td>
<td>2.24</td>
</tr>
<tr>
<td>$M^{CCQE}_A$ (Quasi-Elastic) (GeV)</td>
<td>$0.99^{+0.248}_{-0.149}$</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>$M^{Res}_A$ (Resonance) (GeV)</td>
<td>$1.12 \pm 0.22$</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>NC Normalization</td>
<td>$1.0 \pm 0.2$</td>
<td>1.05</td>
<td>1.06</td>
</tr>
<tr>
<td><strong>Oscillation:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_{13}$ ($^\circ$)</td>
<td>$8.5 \pm 0.21$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\theta_{23}$ ($^\circ$)</td>
<td>-</td>
<td>49.8</td>
<td>50.2</td>
</tr>
<tr>
<td>$\Delta m_{32}^2$ ($10^{-3} eV^2$)</td>
<td>-</td>
<td>2.53</td>
<td>2.56</td>
</tr>
<tr>
<td><strong>Detector:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optical Eff., Overall (%)</td>
<td>$100 \pm 10$</td>
<td>98.4</td>
<td>98.4</td>
</tr>
<tr>
<td>Optical Eff., Lateral ($\sigma$)</td>
<td>$0.0 \pm 1.0$</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td>Optical Eff., Head-on (a.u.)</td>
<td>-</td>
<td>$-0.63$</td>
<td>$-0.64$</td>
</tr>
<tr>
<td>Local Ice Model</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bulk Ice, Scattering (%)</td>
<td>$100.0 \pm 10$</td>
<td>103.0</td>
<td>102.8</td>
</tr>
<tr>
<td>Bulk Ice, Absorption (%)</td>
<td>$100.0 \pm 10$</td>
<td>101.5</td>
<td>101.7</td>
</tr>
<tr>
<td><strong>Atmospheric Muons:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atm. $\mu$ Fraction (%)</td>
<td>-</td>
<td>8.1</td>
<td>8.0</td>
</tr>
<tr>
<td>$\Delta \gamma_\mu$ ($\mu$ Spectral Index, $\sigma$)</td>
<td>$0.0 \pm 1.0$</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Coincident $\nu + \mu$ Fraction</td>
<td>$0.0 + 0.1$</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Measurement:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_\tau$ Normalization</td>
<td>-</td>
<td>0.73</td>
<td>0.57</td>
</tr>
</tbody>
</table>
Background Rates

<table>
<thead>
<tr>
<th>Type</th>
<th>Analysis $A$ Events $\pm 1\sigma$</th>
<th>Analysis $B$ Events $\pm 1\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_e + \bar{\nu}_e$ CC</td>
<td>13462 29</td>
<td>9545 23</td>
</tr>
<tr>
<td>$\nu_e + \bar{\nu}_e$ NC</td>
<td>1096 9</td>
<td>923 8</td>
</tr>
<tr>
<td>$\nu_\mu + \bar{\nu}_\mu$ CC</td>
<td>35706 48</td>
<td>23852 39</td>
</tr>
<tr>
<td>$\nu_\mu + \bar{\nu}_\mu$ NC</td>
<td>4463 19</td>
<td>3368 17</td>
</tr>
<tr>
<td>$\nu_\tau + \bar{\nu}_\tau$ CC</td>
<td>1804 9</td>
<td>934 5</td>
</tr>
<tr>
<td>$\nu_\tau + \bar{\nu}_\tau$ NC</td>
<td>556 3</td>
<td>445 4</td>
</tr>
<tr>
<td>Atmospheric $\mu$</td>
<td>5022 167</td>
<td>1889 45</td>
</tr>
<tr>
<td>Noise Triggers</td>
<td>93 27</td>
<td>&lt; 25 &lt; 5</td>
</tr>
<tr>
<td>total (best fit) observed</td>
<td><strong>62203 180</strong></td>
<td><strong>40959 68</strong></td>
</tr>
<tr>
<td>total observed</td>
<td><strong>62112 249</strong></td>
<td><strong>40902 202</strong></td>
</tr>
</tbody>
</table>
DOM PE Charge

- DOM records a shower of PEs per photoelectron that gets knocked off the photocathode.
- Records pulses as a function of time.
- A discriminator threshold is set to reject noise and accept signals.
- A fit for one PE (SPE) is set and all other pulses are fit compared to that.
- So if a pulse was 3/4 of that, the PE charge would read “0.75 PE charge.”
- The discriminator threshold is set to accept any pulse above 0.25.
# XGBoost BDT Feature Importance

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Feature importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>reconstructed cascade energy</td>
<td>0.081</td>
</tr>
<tr>
<td>reconstructed track length</td>
<td>0.657</td>
</tr>
<tr>
<td>reconstructed $\cos(\theta_{\text{zenith}})$</td>
<td>0.068</td>
</tr>
<tr>
<td>reconstructed $\cos(\theta_{\text{zenith}})$ uncertainty</td>
<td>0.099</td>
</tr>
<tr>
<td>LLH ratio of cascade+track reco to shower only reco</td>
<td>0.096</td>
</tr>
</tbody>
</table>
Label: Track (23.6 GeV)

DOM number

Time (nsec)

Track (23.6 GeV)

Conv2D

MaxPool

DOM axis

Time axis
Label: Cascade (26.3 GeV)

DOM number

Time (nsec)

Conv2D

MaxPool

DOM axis

Time axis
IceCube Upgrade

Ref: Stuttard 20181107