PROCEEDINGS OF SCIENCE

Reactor Neutrino Energy Reconstruction with Machine Learning Techniques for the JUNO Experiment

Arsenii Gavrikov[a,b,*] **on behalf of the JUNO collaboration**

[a] *Dipartimento di Fisica e Astronomia dell'Università di Padova, Via Marzolo 39, Padova, Italia*
[b] *INFN Sezione di Padova, Via Marzolo 39, Padova, Italia*

E-mail: arsenii.gavrikov@pd.infn.it

The Jiangmen Underground Neutrino Observatory (JUNO) with its satellite Taishan Antineutrino Observatory (TAO) is a next-generation neutrino experiment with a broad physics program. Currently under construction, JUNO is expected to start data-taking in 2024. The central detector of JUNO is an acrylic sphere filled with 20 kt of liquid-scintillator (LS) surrounded by 43212 photomultiplier tubes (PMTs).

The primary goals of the experiment are to determine the neutrino mass ordering (NMO) within 3-4σ in 6 years and to measure neutrino oscillation parameters $\sin^2\theta_{12}$, Δm^2_{21}, Δm^2_{31} with sub-percent precision. To achieve the goals, JUNO will study reactor antineutrino emitted from two nuclear power plants located 52.5 km away from the detector.

The main requirement for JUNO is a high energy resolution. The detector is constructed to provide an energy resolution of 3% at 1 MeV. In this study, neutrino energy reconstruction with machine learning techniques is presented. The reconstruction techniques are based on aggregated information collected by PMTs. Two models are considered: Boosted Decision Trees and Fully Connected Deep Neural Network. Moreover, the transferability of the approach is shown with an example of JUNO's satellite detector TAO.

The European Physical Society Conference on High Energy Physics (EPS-HEP2023)
21-25 August 2023
Hamburg, Germany

[*] Speaker

https://pos.sissa.it/

PoS(EPS-HEP2023)189

1. Introduction

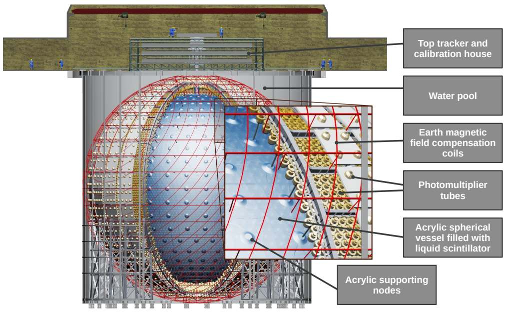

Figure 1: Schematic view of the JUNO central detector and other main components

The Jiangmen Underground Neutrino Observatory (JUNO) is a neutrino observatory with a broad physics program located in China 52.5 km away from two power plants: Taishan and Yangjiang [1]. The central detector (CD) of JUNO is an acrylic sphere 35.4 meters in diameter filled with 20 kt of LS. The CD target is watched by a large number of PMTs of two types: 17612 large 20-inch tubes (LPMTs) and 25600 small 3-inch tubes (SPMTs). The PMT array provides almost 78% of photo coverage of the detector. To provide JUNO with a reference spectrum, a satellite experiment, called Taishan Antineutrino Observatory (TAO) is also under construction [2]. The TAO's detector is an acrylic sphere of 1.8 m diameter, filled with 2.8 tons of gadolinium-doped LS. Being located 30 m away from one of the Taishan reactor cores, TAO will be able to measure the reactor neutrino energy spectrum with a resolution of ~2% at 1 MeV. To achieve this energy resolution, the detector is equipped with 4024 silicon photomultipliers that cover ~94% of the sphere. Figure 2 shows a schematic view of TAO.

The main goal of the JUNO experiment is to determine the neutrino mass ordering (NMO) within 3-4σ in 6 years. The main requirement to ensure resolving NMO is an energy resolution of ~3% at 1 MeV. Moreover, JUNO will be able to measure the following oscillation parameters $\sin^2\theta_{12}$, Δm^2_{21}, Δm^2_{31} with sub-percent precision.

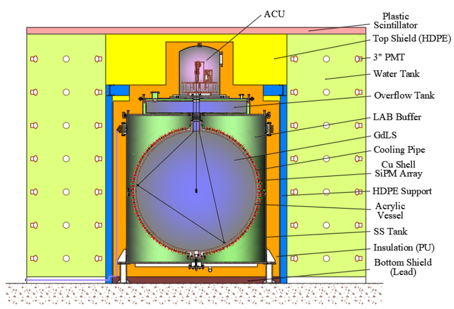

Figure 2: A schematic view of the TAO detector and its main components.

The main reaction for detecting neutrinos in JUNO is Inverse Beta Decay (IBD): $\overline{\nu}_e + p \to e^+ + n$. The PMT array collects optical photons emitted within the signals. The collected information is used to reconstruct the deposited energy analysing the temporal distribution and spatial pattern of fired PMTs. Since almost all the neutrino energy converts into the energy of the positron, by reconstructing its energy we can recover the initial neutrino energy using the following expression: $E_{\tilde{\nu}_e} \approx E_{e^+} + 0.8$ MeV. In this study, we use only positron events, assuming event selection is performed.

2. Machine Learning Approach

As previously mentioned, the JUNO central detector is equipped with a huge number of PMTs: 17612 large PMTs and 25600 small ones. From each PMT two values can be extracted: the charge and first hit time (FHT). Thus, each event is described by 86424 variables (many of them are usually zeros). Such a high-dimensional input requires additional processing. By increasing the complexity of models and reducing the leakage of information, the following sequence of approaches can be considered: 1) an approach based on aggregated information with simple models, 2) convolutional neural networks, and 3) graph neural networks.

In this work, the aggregated features approach is studied. The aggregated features are used as input for the following models: Boosted Decision Trees (BDT) and Fully Connected Deep Neural Network (FCDNN). The study is a continuation of Ref. [3] and uses an updated JUNO software [4] and also adds information collected by SPMTs. Moreover, the transferability of the approach to other LS-based detectors is described with the example of the JUNO-TAO detector (Section 2.2). The study of PMT-wise approaches 2) and 3) can be found in Ref. [5].

The advantage of the simpler approach 1) is that it is more interpretable since the aggregated features are constructed based on physical intuition and exploratory data analysis. Moreover, a small number of features provides faster inference. On the other hand, approaches 2) and 3) lose less information and can potentially achieve higher precision. The drawbacks of using more complex models are limited interpretability, slower inference, and more complicated training. The latter can be especially important in the early phase of data-taking when models need to be retrained every time according to the constantly updated simulation software.

Notation	*Description*
`AccumCharge`	Total accumulated charge
`nPMTs`	Number of fired PMTs
x_{cc}, y_{cc}, z_{cc}, R_{cc}, θ_{cc}, ϕ_{cc}, J_{cc}, ρ_{cc}, γ_z^{cc}, γ_y^{cc}, γ_x^{cc}	Center of charge
$\mathrm{pe}_{\mathrm{n\%}}$, $\mathrm{pe}_{\mathrm{mean}}$, $\mathrm{pe}_{\mathrm{std}}$, $\mathrm{pe}_{\mathrm{skew}}$, $\mathrm{pe}_{\mathrm{kurtosis}}$	Charge distribuion
x_{cht}, y_{cht}, z_{cht}, R_{cht}, θ_{cht}, ϕ_{cht}, J_{cht}, ρ_{cht}, γ_z^{cht}, γ_y^{cht}, γ_x^{cht}	Center of FHT
$\mathrm{ht}_{\mathrm{n\%}}$, $\mathrm{ht}_{\mathrm{n_{i+1}\%-n_i\%}}$, $\mathrm{ht}_{\mathrm{mean}}$, $\mathrm{ht}_{\mathrm{std}}$, $\mathrm{ht}_{\mathrm{skew}}$, $\mathrm{ht}_{\mathrm{kurtosis}}$	FHT distribution

Table 1: List of all feature notations with brief descriptions. Here, $\mathrm{n} \in \{2, 5, 10, 15, ..., 90, 95\}$.

Table 1 collects and briefly describes the full set of engineered features and the detailed description can be found in Ref. [3]. `AccumCharge` is the total charge accumulated on all fired PMTs and it is at first order proportional to E_{dep}. Other features, e.g., the center of charge and center of FHT provide a rough approximation of the vertex and help to correct the non-uniformity of the response. Additional information about the signal is extracted from charge and FHT distributions by computing their moments and decomposing them with percentiles. In total, the set consists of 91 features.

2.1 JUNO central detector

To train models and to evaluate their performance, we prepared the training and the testing datasets, generated by the full detector Monte Carlo method using the official JUNO software [4].

Many of the 91 features, described in the previous section, are highly correlated and we want to keep only a subset of them which provides the same performance of the model as the full set. By using a greedy algorithm for feature selection described in Ref. [3], we kept the following 17 features (sorted by importance):

$$\texttt{AccumCharge}, R_{\mathrm{cht}}, \mathrm{J_{cc}}, \mathrm{ht}_{20\%-15\%}, \mathrm{pe}_{\mathrm{std}}, \mathrm{nPMTs}, \mathrm{z_{cc}}, \mathrm{ht}_{\mathrm{std}}, \mathrm{R_{cc}}, \mathrm{ht}_{30\%-25\%},$$
$$\mathrm{ht}_{5\%-2\%}, \mathrm{pe}_{\mathrm{mean}}, \mathrm{ht}_{15\%-10\%}, \mathrm{ht}_{25\%-20\%}, \mathrm{ht}_{35\%-30\%}, \mathrm{ht}_{10\%-5\%}, \mathrm{pe}_{50\%}.$$

2.2 TAO detector

To test the transferability of the approach for other LS-based detectors, we used the JUNO's satellite near detector TAO. Using analogue datasets generated for the TAO detector (~2M for training), we performed the same procedures of the feature engineering and feature selection. The following optimized set of 12 features was used to train the models (sorted by importance):

$$\texttt{AccumCharge}, \rho_{\rm cc}, {\rm ht}_{35\%}, {\rm pe}_{90\%}, {\rm pe}_{\rm mean}, \texttt{nSiPMs}, {\rm ht}_{5\%}, {\rm R}_{\rm cc}, {\rm pe}_{\rm std}, {\rm ht}_{75\%}, {\rm pe}_{\rm kurtosis}, {\rm ht}_{15\%}.$$

Note that, unlike the set for the JUNO's detector, charge-related features are more dominant in the set for TAO. Temporal information becomes less important for TAO due to its smaller size.

3. Results and conclusions

Figure 3 shows the energy reconstruction performance (resolution and bias) of the BDT and FCDNN models. The hyperparameter optimization of the models is performed and follows the same procedure as in Ref [3].

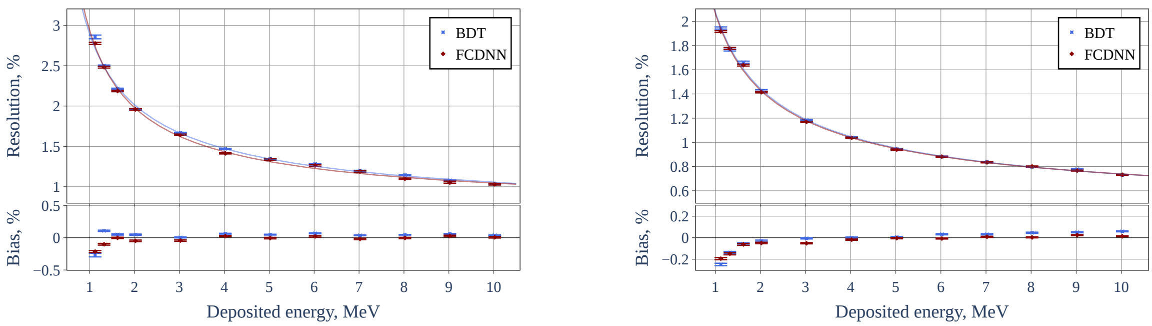

Figure 3: Energy reconstruction performance for BDT and FCDNN: a) JUNO detector and b) TAO detector.

In this study, we present an application of machine learning techniques for precise energy reconstruction for the JUNO detector and its satellite detector TAO. We use two models, BDT and FCDNNN, trained using aggregated features extracted from simulation data.

Acknowledgments

We are very thankful to Yury Malyshkin and Fedor Ratnikov for contributing to this work. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101034319 and from the European Union – NextGenerationEU.

References

[1] A. Abusleme *et al.*, JUNO physics and detector, PPNP **123**, 103927 (2022).
[2] A. Abusleme *et al.*, TAO Conceptual Design Report, arXiv:2005.0874 (2020).
[3] A. Gavrikov et al., Energy reconstruction for large liquid scintillator detectors with machine learning techniques: aggregated features approach, EPJ C **82**, 1021 (2022).
[4] T. Lin *et al.*, Simulation software of the JUNO experiment, EPJ C **83**, 382 (2023).
[5] Z. Qian, *et al.*, Vertex and energy reconstruction in JUNO with machine learning methods, NIM-A **1010**, 165527 (2021).