Investigating the Use of Omnifold in Unfolding Jet Transverse Momenta

Simon Browning of Abilene Christian University for the STAR Collaboration

ABILENE CHRISTIAN UNIVERSITY

www.acu.edu

Abstract

A common issue in experimental particle physics is the presence of detector effects that distort the information measured by the detector, e.g., finite resolution smearing energy and momentum measurements. Thus, it is necessary to create and implement techniques that correct for the detector effects. These effects become especially challenging when the underlying spectra are steeply falling, e.g., momentum distributions of hadrons within jets. Iterative Bayesian Unfolding (IBU) is a well-established technique, using properties of Bayesian statistics to predict the true particle data based on the detector data. IBU requires binning the data, but Omnifold, a new algorithm created in 2020, implements a machine learning approach that enables IBU event-by-event. We train a Keras classifier on fast Monte Carlo data, which we generate using functional distributions to simulate the steeply falling true spectra and the smearing from detector responses of the STAR detector at RHIC. In this study, we investigate the ability of Omnifold to recover the true distributions from different starting assumptions. This is relevant for measurements such as the distributions of hadrons within jets in the STAR 2017 510 GeV *p+p* data set. A status report will be presented.

The STAR Experiment

Brookhaven National Lab (BNL) studies the strong interaction through collisions between subatomic particles. The STAR experiment is housed at BNL and focuses on studying the quark-gluon plasma (QGP) and spin structure through polarized *p+p* collisions. The STAR experiment utilizes the STAR detector, consisting of a time projection chamber (TPC), electromagnetic calorimeter (ECal) and hadronic calorimeter (HCal), to make these studies possible.

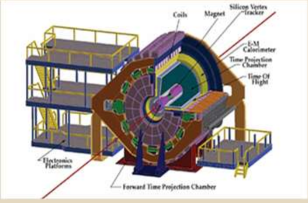


Fig. 1: The STAR Experiment.

What is Omnifold?

Omnifold is a machine-learning-based algorithm that can unfold multiple unbinned observables simultaneously. The algorithm deals with 2 pairs of data sets: a natural set and a synthetic, Monte Carlo set. Within each pair, there is a detector data set and a true data set, where the detector data is the smeared version of the true data. Omnifold works in 2 steps. Step 1 reweights the synthetic detector set to match the natural detector set, giving us the pull weights. Step 2 uses the synthetic true set to convert the pull weights into push weights that can be used to reweight the synthetic true set. The reweighted synthetic true set gives the distribution that is most likely to have produced the nature detector set, ie Omnifold's estimate of the nature true set. We use the Keras neural network classifier for each step, consisting of 3 hidden layers and 50 nodes per hidden layer. For testing, we generate a pair of Monte Carlo synthetic data sets and a pair of Monte Carlo natural data sets, then compare Omnifold's prediction to the natural true particle data.

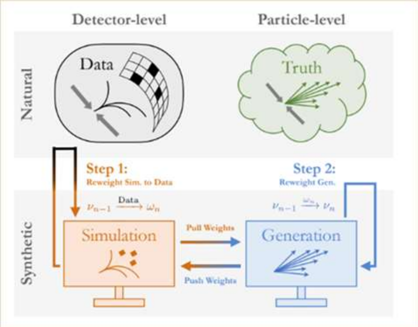
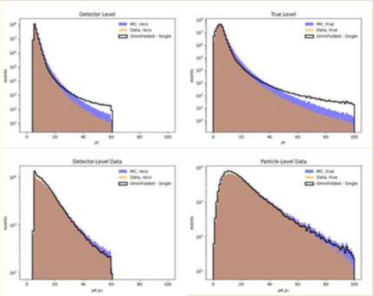


Fig. 2: The four data sets needed for Omnifold and how the sets relate to each other.
Image Source: Andreassen, Anders, et al. *OmniFold: A Method to Simultaneously Unfold All Observables*. 2019.

Steeply Falling Spectra

Omnifold has not been extensively tested on steeply falling spectra. While Omnifold works well on datasets whose largest and smallest bins differ by no more than a few orders of magnitude, we test it on steeply falling spectra typically ranging over 7 to 8 orders of magnitude. Omnifold reliably fits the peaks of the spectra but struggles with fitting the tails.

Steeply Falling Spectra



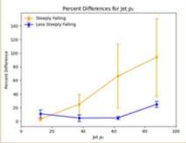


Fig. 3: Uppermost plots show result of using the unaltered Omnifold algorithm on a steeply falling spectrum. Middle plots show result of same procedure done on a less steep spectrum. Final plot shows average percent differences for multiple tests. We use Root to generate Monte Carlo data based on advanced simulations of the STAR detector response.

Custom Weights

We altered Omnifold by giving the classifier for each of the two steps weights that had been changed by a custom weighting function. This function was chosen to rescale the initial weights such that the tail of the steeply falling spectra would have a similar value to the peak.

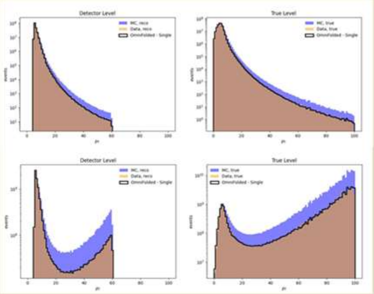


Fig. 4: Uppermost plots show result of using Omnifold with the custom reweighting function. Middle plots show what Omnifold "sees" when fitting. The custom reweighting function used here is $Weight_{new} = Weight_{old}\exp(p_T^{0.68})$.

Results

Omnifold consistently fits steeply falling spectra more effectively with the implementation of a custom reweighting function. Not only is the final fit more accurate than before, but the altered Omnifold algorithm fits more consistently as well. This results in the fits having a lower average percent error as well as lower standard deviations than before.

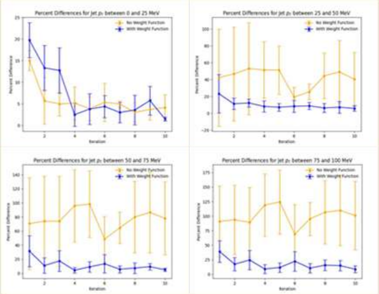


Fig. 5: Average percent difference between Omnifold fit and natural true data, with and without the custom reweighting function. Each plot shows results from different 25 MeV ranges.

Future Work

Our custom reweighting function was chosen due to our steeply falling spectra decaying roughly exponentially. However, we would like a reproducible way of fitting a function to our spectra and using that to create a custom reweighting function. We believe that an ideal reweighting function would flatten the curve that Omnifold "sees". A fit of this nature has not yet been achieved by us, as the form of the function is likely complicated.

Acknowledgements

Special thanks to:

- Dr. Joseph Atchison
- Dr. Jim Drachenberg
- The Department of Energy
- The ACU Nuclear Physics Research Group

BROOKHAVEN NATIONAL LABORATORY

STAR

This research was supported by US DOE MENP Grant DE-FG02-03ER41243.

DEPARTMENT OF ENGINEERING AND PHYSICS