Investigation of the Fast Boosted Decision Tree for B Meson Classification

Saleh Naghdi

Supervised by Martin Sevior

THE UNIVERSITY OF MELBOURNE

THE LABY FOUNDATION

Introduction

Violation of CP symmetry in certain particle decay chains has become of great interest to particle physicists for its underpinning of unexplored and new physics. One such decay channel currently being investigated by the Belle II Experiment is that of the neutral B meson:

$$B^0 \rightarrow K^+ K^- \qquad \textit{Equation 1}$$

As well as producing the *signal* event in Equation 1, the parent interaction e^+e^- can generate other *background* events according to the pathway outlined in Figure 1, where *generic* and *continuum* events collectively comprise the background.

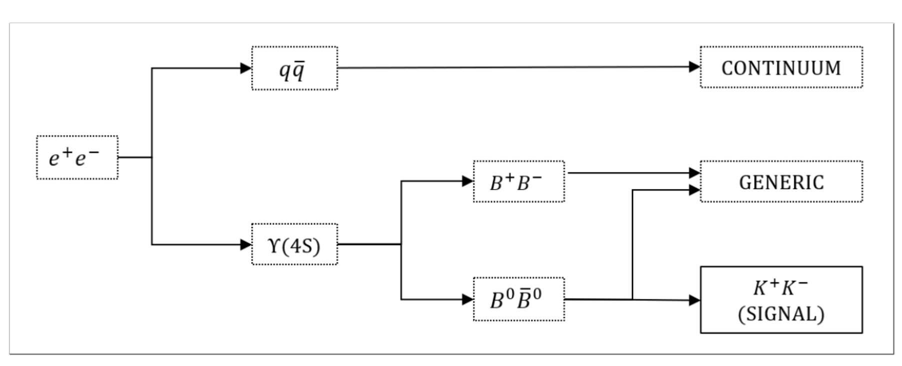


Figure 1: Summary of e^+e^- decay channels at the Belle II experiment.

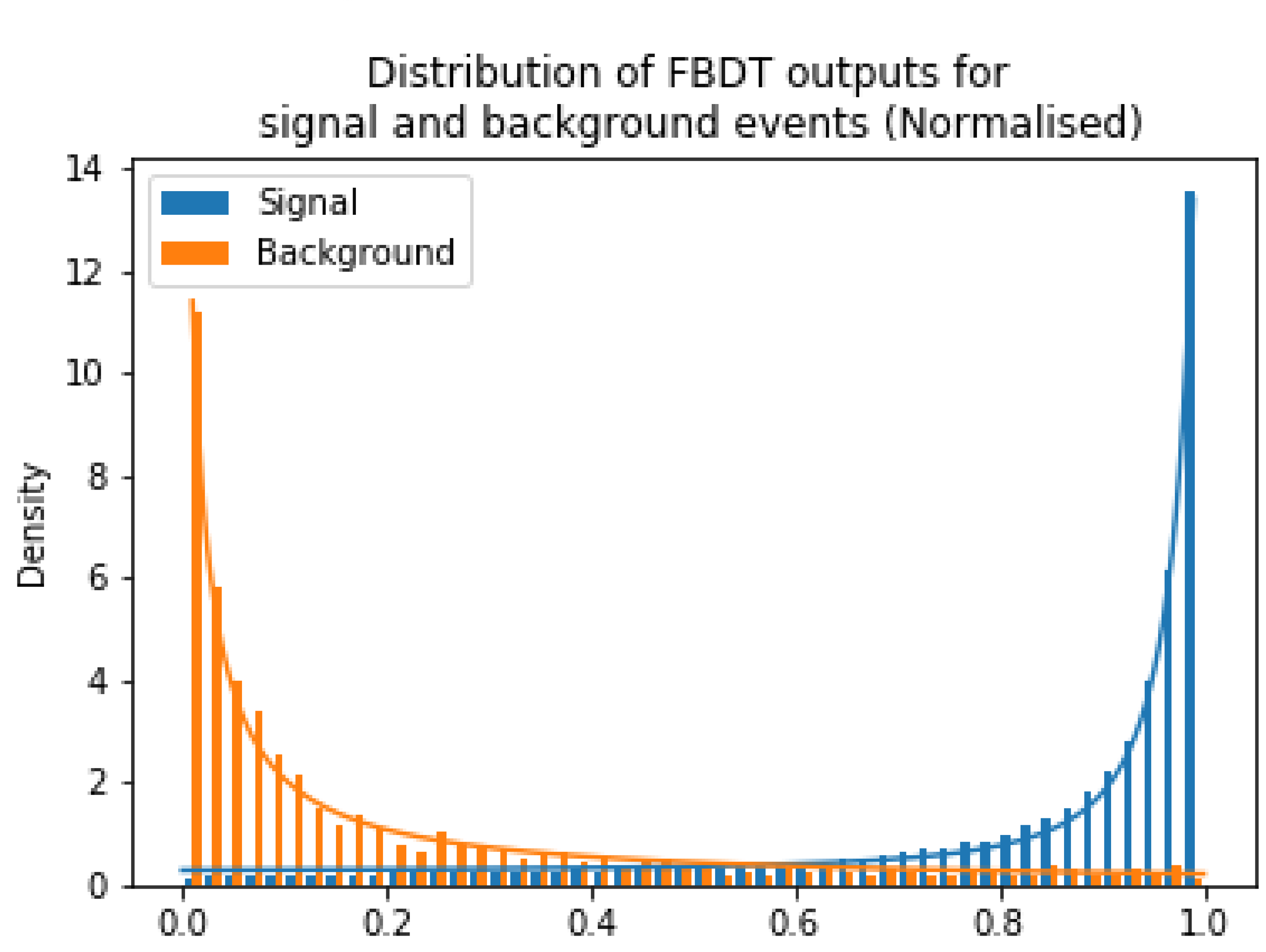
Often, filtering signal events from background can be difficult. The Fast Boosted Decision Tree (FBDT) is one possible background suppression technique [1] that outputs the likelihood of an event being a signal, and it may be trained using simulated labeled data from the Belle II experiment. This project sets out to investigate further refinements that could be made in both the training data and evaluative metrics used, in hopes of optimizing background suppression.

Method

One measure of classification accuracy is the *Figure of Merit* (FOM), which, for a trained model that admits a total of S signal events (true positives) and B background events (false positives), is defined by

$$FOM = \frac{S}{\sqrt{B+S}}. \qquad \textit{Equation 2}$$

A trained FBDT assigns a significance value to each event (see graph 1). Provided a *cut-off* value, only events with significances above the cut are retained (and thus classified to be signal) while the rest are suppressed as background. By floating the cut value, the maximum FOM attained can be read as an indicator of a significant measurement



Graph 1

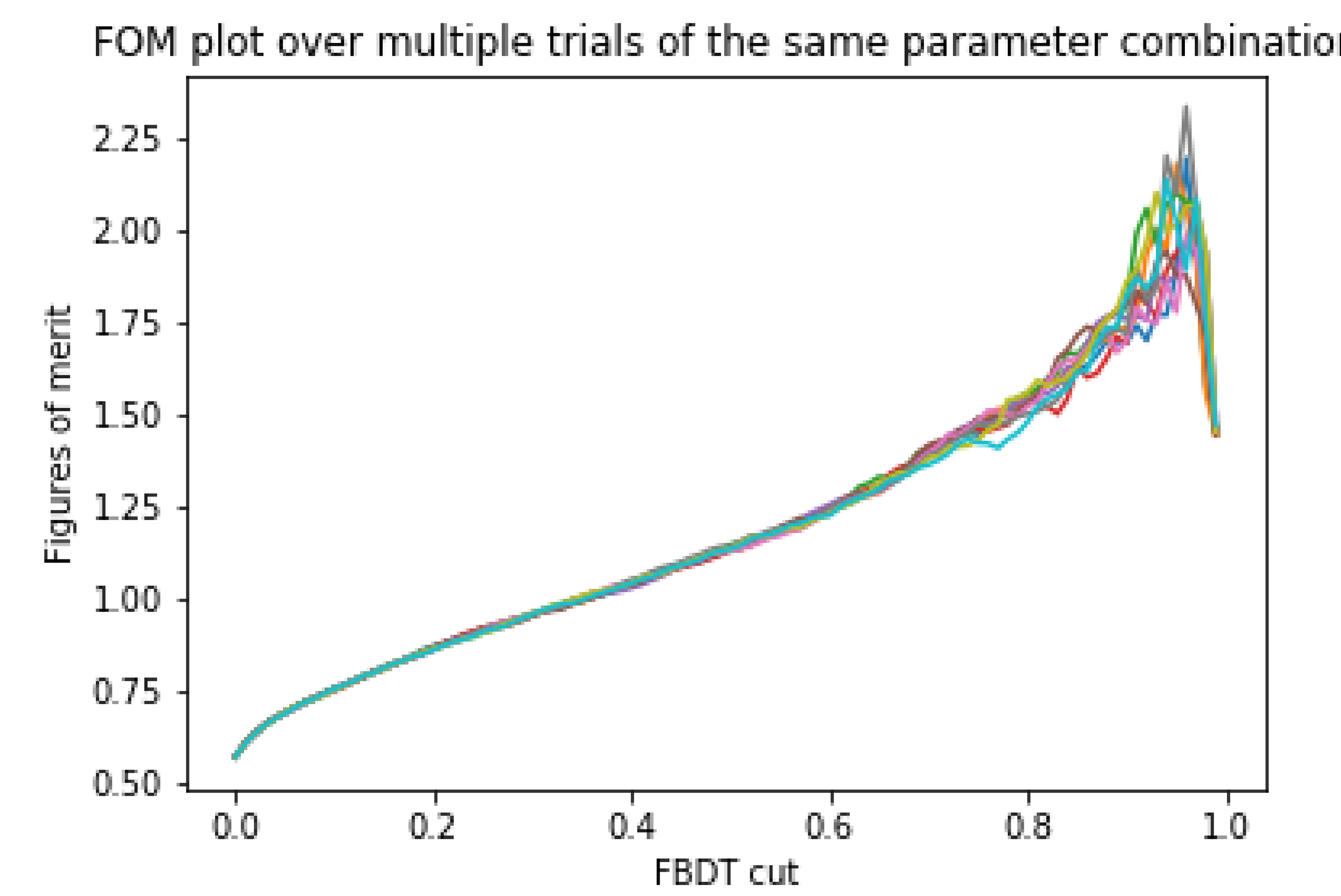
Batch processing was used to train multiple FBDTs conditioned on the same dataset and hyperparameter settings. After testing, the variance and average of the maximum FOM was calculated and a comparison of classification accuracy between different training datasets was conducted.

Additionally, training was done on a new Particle Identification (PID) dataset which contained additional information about the trajectory and identity of charged daughter particles produced in the decay channels. An engineered PID dataset was also constructed, which subsumed the five particle identification variables into a single *TrkMax* variable defined by

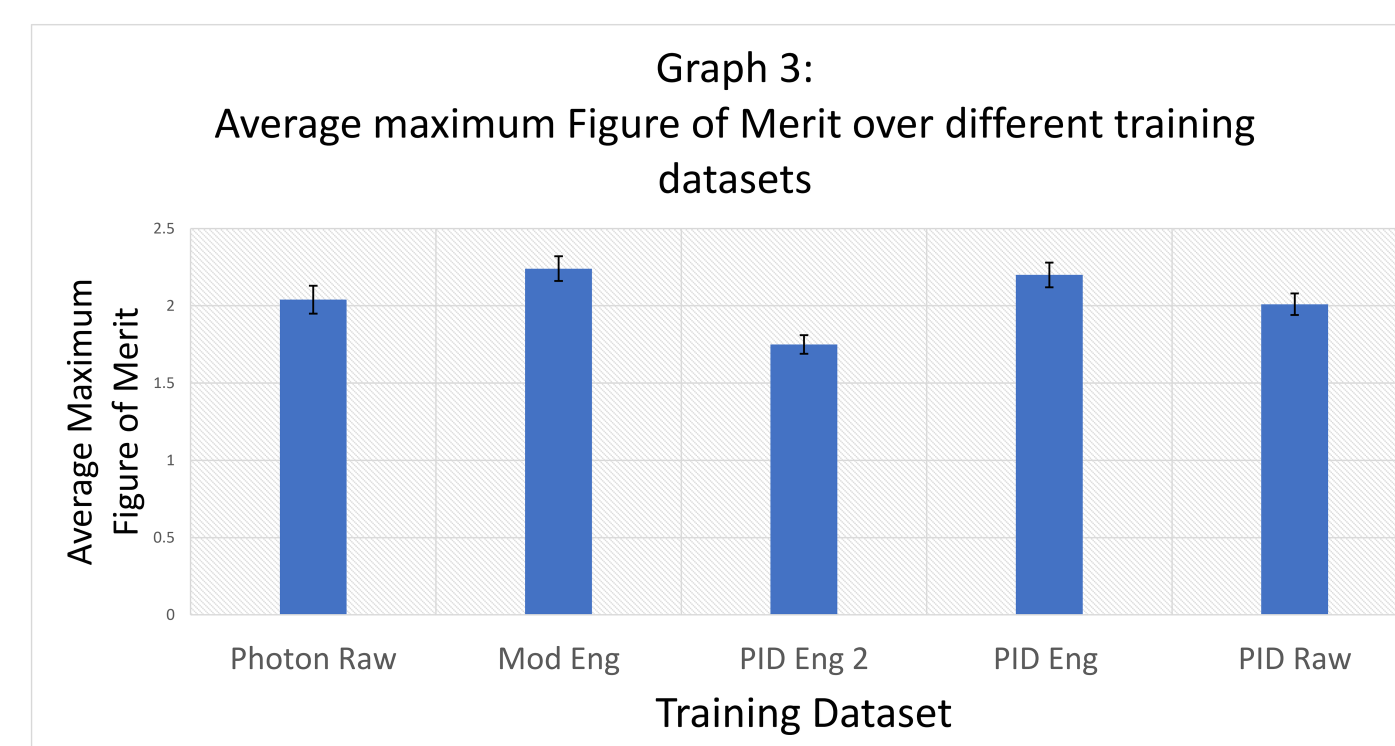
$$TrkMax = \{particle \mid p_{particle} = \max(p_{e^-}, p_K, p_\pi, p_\mu, p_p)\}$$

The optimum hyperparameters resulting in the best maximum FOM were then deduced using the same simplified grid search method employed in [2].

Results



***Graph 2**: FOM plots of 10 identical trials (overlaid in different colours)*



***Graph 3**: Datasets Photon Raw and Mod Eng correspond to sets D and B from [2] respectively. As introduced in this project, we also have the new Particle Identification dataset (PID Raw), and its feature engineered version (PID ENG) as described in the Method. PID Eng 2 further builds on PID Eng by removing redundant variables.*

Discussion

The variation in the maximum FOM seen in graph 2 suggests the presence of randomization within the training process of the FBDT that is irrespective of the choice of training dataset or hyperparameters. The fact that this variation is most salient at around the maximum also suggests this effect is most sensitive on the best FOM attained.

It is thus unreliable to compare single instances – and not averages – of differentially parameterized FBDTs, which is what was done in the greedy method used in [2]. Fortunately, the standard deviation remained at a consistent value for samples larger than 10 trials, suggesting a minimum trial sample size for which means may be reliably compared.

A comparison of the mean maximum FOMs over various training datasets is shown in graph 3. As was found in [2], the engineered datasets perform better than their raw alternatives.

While PID Eng (our dataset) and Mod Eng (from [2]) are similar in performance, this comparison may be unreliable as the reduced grid search method of hyperparameter optimisation used for each dataset compared only FBDT instances and not averages of them. Moreover, this method assumes the independence of different hyperparameters, which need not be the case. More valid methods could involve evaluating all possible hyperparameter combinations, which, while computationally demanding, spares the assumption of independence.

Conclusion

Where concerning evaluative methods, variation in the maximum FOM attained by the same FBDT suggests that an averaging over multiple trials is required before comparisons can be made between different FBDT instances.

The FBDT trained on the raw PID dataset was found to perform just as well as the other raw datasets used previously, while the engineered version is comparable to the best dataset, being Mod Eng.

References

[1]: Keck, T., 2017. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Computing and Software for Big Science*, 1(1).

[2]: Porter, D., 2020. Investigation of a machine learning algorithm for analysis of particle physics data.