Separation of $\bar{c}c, \bar{b}b$ using ML

ALEXANDRE BEAUBIEN, PHD STUDENT

UNIVERSITY OF VICTORIA

THE BELLE II COLLABORATION

University of Victoria

2022-06-07

Outline

1. Motivation
2. Model
3. Results

Motivation

We want to be able to seperate $c\bar{c}$ and $b\bar{b}$ events with high efficiency for A_{LR} measurements.

Build ML based classifier (gradient boosted decision tree – GBDT) using **event shape variables.**

Results:

Sufficient for selecting $b\bar{b}$ events, but further selections are needed to select $c\bar{c}$ events.

BDT model

Evaluation variables:

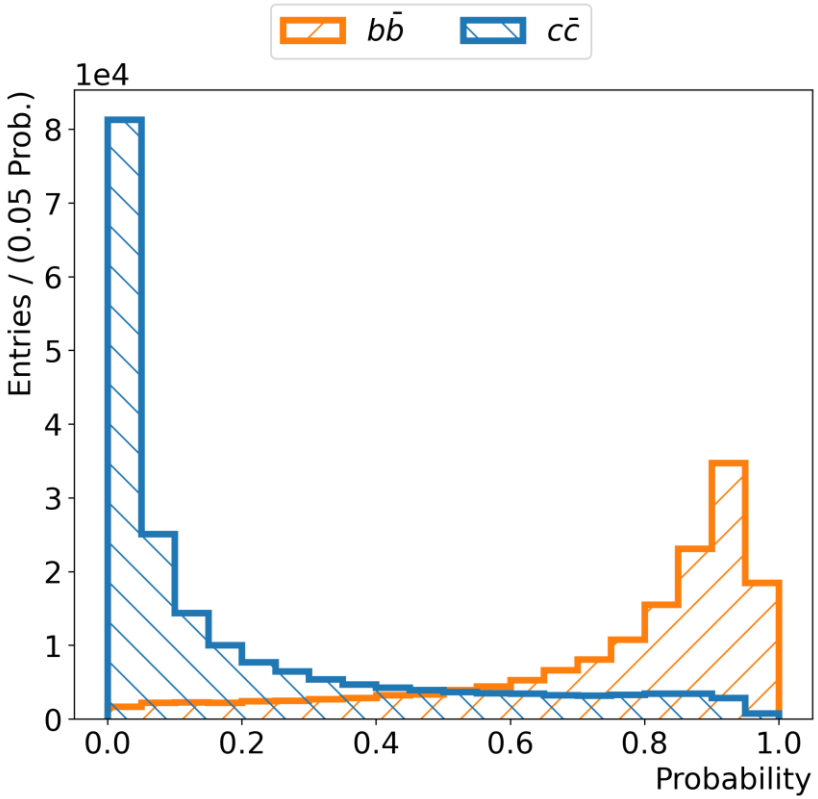
$$b\bar{b}\text{ Efficiency} = \frac{\#\text{ of real } b\bar{b} \text{ predicted as } b\bar{b}}{\#\text{total number of MC generated } b\bar{b}}$$

$$c\bar{c}\text{ Fake rate} = \frac{\#\text{ of real } c\bar{c} \text{ misidentified as } b\bar{b}}{\#\text{total number of MC generated } c\bar{c}}$$

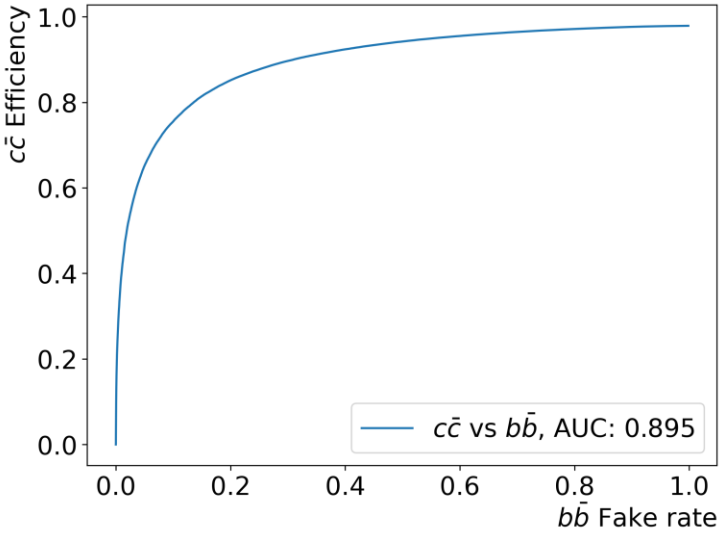
Variable	Feature Importance
foxWolframR2	0.590
thrust	0.184
foxWolframR1	0.081
harmonicMomentThrust0	0.060
thrustAxisCosTheta	0.039
harmonicMomentCollision2	0.020
foxWolframR3	0.010
aplanarity	0.006
harmonicMomentThrust2	0.006
sphericity	0.004

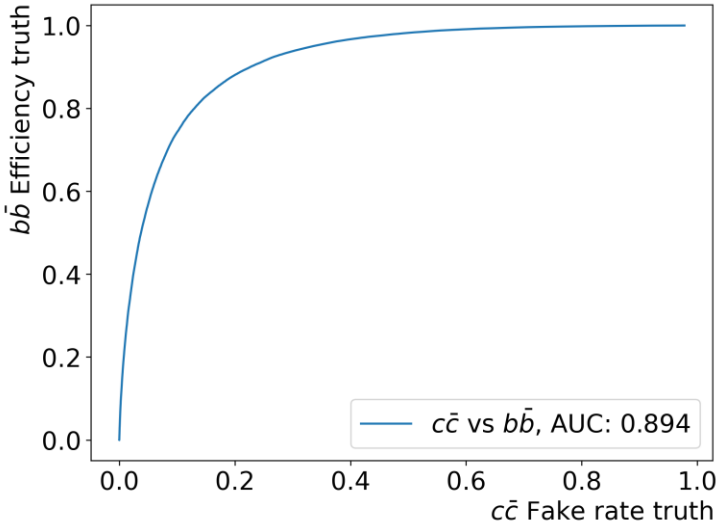
10 most important variables (as defined by the GBDT)

Results ROC curves



Use **ROC** and **AUC** to quantify model quality. Rates are normalized to # of generated particles (MC number).





Results $c\bar{c}$

$c\bar{c}$ selection against bb

Fraction	
$c\bar{c}$ efficiency	$b\bar{b}$ fake rate
0.102	0.000
0.201	0.002
0.301	0.005
0.402	0.010
0.500	0.020
0.599	0.037
0.700	0.069
0.800	0.138
0.900	0.311
0.979	0.998

(a) Efficiency and fake rates.

Events per nb^{-1}	
$c\bar{c}$	$b\bar{b}$
0.133	0.000
0.261	0.002
0.391	0.006
0.523	0.011
0.650	0.022
0.779	0.041
0.910	0.077
1.040	0.153
1.170	0.345
1.273	1.108

(b) Number of events selected as $c\bar{c}$..

Table 4: Classification of $c\bar{c}$ events against $b\bar{b}$ events. These tables show the results from figure 7a.

Results $b\bar{b}$

$b\bar{b}$ selection against $c\bar{c}$ and uds(tau) background in fraction

Fraction						
Efficiency $b\bar{b}$	Background Fraction	Fake Rate $c\bar{c}$	$u\bar{u}$	$d\bar{d}$	$s\bar{s}$	$\tau^+\tau^-$
0.098	0.088	0.003	0.002	0.002	0.002	0.002
0.196	0.091	0.008	0.004	0.004	0.004	0.002
0.301	0.110	0.015	0.008	0.008	0.008	0.003
0.401	0.128	0.024	0.013	0.013	0.014	0.003
0.501	0.153	0.038	0.020	0.020	0.022	0.003
0.599	0.181	0.056	0.029	0.030	0.032	0.004
0.699	0.221	0.083	0.044	0.046	0.050	0.004
0.800	0.281	0.130	0.070	0.073	0.079	0.006
0.900	0.380	0.226	0.124	0.131	0.144	0.014
0.998	0.802	0.979	0.966	0.966	0.941	0.996

Results $b\bar{b}$

$b\bar{b}$ selection against $c\bar{c}$ and uds(tau) background in nb

Events selected as $b\bar{b}$ per nb^{-1}					
$b\bar{b}$	$c\bar{c}$	$u\bar{u}$	$d\bar{d}$	$s\bar{s}$	$\tau^+\tau^-$
0.109	0.004	0.003	0.001	0.001	0.002
0.218	0.010	0.006	0.002	0.002	0.002
0.334	0.020	0.013	0.003	0.003	0.003
0.445	0.031	0.021	0.005	0.005	0.003
0.556	0.049	0.032	0.008	0.008	0.003
0.665	0.073	0.047	0.012	0.012	0.004
0.776	0.108	0.071	0.018	0.019	0.004
0.888	0.169	0.113	0.029	0.030	0.006
0.999	0.294	0.200	0.052	0.055	0.013
1.108	1.273	1.555	0.386	0.358	0.915

Lepton selection

Include the requirement that the **event must have a lepton**. Muon selection is not optimal. Does reduce the uds contribution. cc and tau not very affected

Fraction						
	MC Truth			ID Cuts		
Type	μ	e	Total	μ	e	Total
$b\bar{b}$	0.26	0.27	0.40	0.33	0.22	0.41
$c\bar{c}$	0.13	0.16	0.28	0.25	0.12	0.34
$u\bar{u}$	0.02	0.06	0.08	0.16	0.05	0.20
$d\bar{d}$	0.02	0.06	0.08	0.16	0.05	0.20
$s\bar{s}$	0.02	0.05	0.07	0.14	0.04	0.17
$\tau^+\tau^-$	0.25	0.27	0.47	0.27	0.21	0.45

(a) Fraction of events containing at least one lepton.

Events per nb^{-1}						
	MC Truth			ID Cuts		
Type	μ	e	Total	μ	e	Total
$b\bar{b}$	0.29	0.30	0.45	0.36	0.24	0.46
$c\bar{c}$	0.17	0.21	0.36	0.32	0.15	0.44
$u\bar{u}$	0.04	0.10	0.13	0.25	0.07	0.31
$d\bar{d}$	0.01	0.02	0.03	0.06	0.02	0.08
$s\bar{s}$	0.01	0.02	0.03	0.05	0.01	0.07
$\tau^+\tau^-$	0.23	0.24	0.44	0.25	0.20	0.41

(b) Number of events containing at least one lepton.

Table 7: Simulation sets containing at least one lepton. The sets are selected using cuts: truth is the number MC generated number of events with leptons. ID cuts uses a cut on the default particle identification tool (>0.95) and on E/p (>0.85 electron).

Lepton selection

Include the requirement that the **event must have a lepton**. Muon selection is not optimal. Does reduce the uds contribution. cc and tau not very affected

Fraction						
	MC Truth			ID Cuts		
Type	μ	e	Total	μ	e	Total
$b\bar{b}$	0.26	0.27	0.40	0.33	0.22	0.41
$c\bar{c}$	0.13	0.16	0.28	0.25	0.12	0.34
$u\bar{u}$	0.02	0.06	0.08	0.16	0.05	0.20
$d\bar{d}$	0.02	0.06	0.08	0.16	0.05	0.20
$s\bar{s}$	0.02	0.05	0.07	0.14	0.04	0.17
$\tau^+\tau^-$	0.25	0.27	0.47	0.27	0.21	0.45

(a) Fraction of events containing at least one lepton.

Events per nb^{-1}						
	MC Truth			ID Cuts		
Type	μ	e	Total	μ	e	Total
$b\bar{b}$	0.29	0.30	0.45	0.36	0.24	0.46
$c\bar{c}$	0.17	0.21	0.36	0.32	0.15	0.44
$u\bar{u}$	0.04	0.10	0.13	0.25	0.07	0.31
$d\bar{d}$	0.01	0.02	0.03	0.06	0.02	0.08
$s\bar{s}$	0.01	0.02	0.03	0.05	0.01	0.07
$\tau^+\tau^-$	0.23	0.24	0.44	0.25	0.20	0.41

(b) Number of events containing at least one lepton.

Table 7: Simulation sets containing at least one lepton. The sets are selected using cuts: truth is the number MC generated number of events with leptons. ID cuts uses a cut on the default particle identification tool (>0.95) and on E/p (>0.85 electron).

Selecting with Kaon

Fraction

	Truth				ID Cuts			
Type	μ	e	K	Total	μ	e	K	Total
$b\bar{b}$	0.177	0.184	0.409	0.244	0.227	0.145	0.408	0.255
$c\bar{c}$	0.077	0.095	0.409	0.129	0.147	0.068	0.392	0.155
$u\bar{u}$	0.008	0.021	0.150	0.022	0.057	0.016	0.159	0.051
$d\bar{d}$	0.008	0.020	0.134	0.020	0.053	0.015	0.145	0.047
$s\bar{s}$	0.012	0.027	0.290	0.037	0.085	0.021	0.271	0.082
$\tau^+\tau^-$	0.015	0.017	0.012	0.006	0.018	0.013	0.021	0.010

Events per nb^{-1}

	Truth				ID Cuts			
Type	μ	e	K	Total	μ	e	K	Total
$b\bar{b}$	0.197	0.204	0.454	0.271	0.252	0.161	0.452	0.283
$c\bar{c}$	0.100	0.123	0.531	0.168	0.191	0.088	0.509	0.201
$u\bar{u}$	0.013	0.034	0.242	0.036	0.092	0.026	0.256	0.083
$d\bar{d}$	0.003	0.008	0.054	0.008	0.021	0.006	0.058	0.019
$s\bar{s}$	0.005	0.010	0.110	0.014	0.032	0.008	0.103	0.031
$\tau^+\tau^-$	0.014	0.016	0.011	0.005	0.017	0.012	0.019	0.009

Selection that has:

- Either muon or electron
- Kaon

Does help reduce the uds background

EARLY RESULTS

Conclusion

$c\bar{c}$ and $b\bar{b}$ classification can be achieved with good quality using only event shape variables.

Some additional selections such as: requiring lepton and or kaon in the event are needed to reduce backgrounds usd(tau) in the $c\bar{c}$ selection.

Might need D meson tagging.