The Belle II flavor tagger

Fernando Abudinén

July 11, 2018

1 Introduction
2 The category-based flavor tagger
3 The deep-neural flavor tagger
4 Tests with Belle data
5 Conclusions and Outlook
Time-dep. CP-Analysis at B-Factories

- $\Upsilon(4S)$ above $B \bar{B}$ prod. threshold
- $\Upsilon(4S) \rightarrow B \bar{B} > 96\%$
- $\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$
  \[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

\[ \Upsilon(4S) \rightarrow B \bar{B} > 96\% \]

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]

$\Upsilon(4S)$ above $B \bar{B}$ prod. threshold

$\Upsilon(4S) \rightarrow B \bar{B} > 96\%$

$\frac{\Gamma(B^+ B^-)}{\Gamma(B^0 \bar{B}^0)} \sim 1.06$

\[ \Rightarrow B\text{-Factory} \]
Category-based flavor tagger

<table>
<thead>
<tr>
<th>Categories</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>$e^-$, $e^+$</td>
</tr>
<tr>
<td>Intermediate Electron</td>
<td></td>
</tr>
<tr>
<td>Muon</td>
<td>$\mu^-$, $\mu^+$</td>
</tr>
<tr>
<td>Intermediate Muon</td>
<td></td>
</tr>
<tr>
<td>KinLepton</td>
<td>$e^-$, $\ell^+$</td>
</tr>
<tr>
<td>Intermediate KinLepton</td>
<td></td>
</tr>
<tr>
<td>Kaon</td>
<td>$K^-$</td>
</tr>
<tr>
<td>KaonPion</td>
<td>$K^-$, $\pi^+$</td>
</tr>
<tr>
<td>SlowPion</td>
<td>$\pi^+$</td>
</tr>
<tr>
<td>FastHadron</td>
<td>$\pi^-$, $K^-$</td>
</tr>
<tr>
<td>MaximumP</td>
<td>$\ell^-$, $\pi^-$</td>
</tr>
<tr>
<td>FSC</td>
<td>$\ell^-$, $\pi^+$</td>
</tr>
<tr>
<td>Lambda</td>
<td>$\Lambda$</td>
</tr>
</tbody>
</table>

Total = 13
Starting Info: Objects in the **tag side** (rest of event).

Magenta Boxes: Multi-variate Methods.

**Default**: Belle II’s Fast-boosted decision tree.  
**Cross check**: 3-layer Perceptron (Only Combiner)
Tagging variables

Two types:
- Particle identification (PID)
- Kinematic
**PID:** Likelihoods combining info from different subdetectors.

**Kin.:** Simple: Momentum, transverse momentum, impact params., polar angle.
Elaborated: recoil mass, energy in $W$ boson direction, miss. momentum., cosines to thrust axis, and others.


⇒ Optimized for CPU: Each Variable is calculated only once for each particle list! (108 instead of 220 calculations).
Category-based procedure

Training: Signal MC \( \Upsilon(4S) \rightarrow B^0_1 (\rightarrow J/\Psi K^0_S) \) \( B^0_2 (\rightarrow \text{generic}) \)
Category-based tagger with Belle II MC

Output

\[ \varepsilon_{\text{Eff}} = 37.2\% \]
\[ \Delta \varepsilon_{\text{Eff}} = 0.3\% \]

Calibration

\[ r_{\text{MC}} = 1 - 2 \cdot w_{\text{MC}} \]

\[ \langle r \rangle_{\text{FBDT}} \approx \frac{1}{\sqrt{\varepsilon_{\text{Eff}} \cdot N}} = \frac{1}{\sqrt{N_{\text{Eff}}}} \]
Category-based tagger with Belle MC

Output

\[ \epsilon_{\text{Eff}} = 34.3\% \]
\[ \Delta \epsilon_{\text{Eff}} = -0.1\% \]

Calibration

\[ r_{\text{MC}} = 1 - 2 \cdot w_{\text{MC}} \]

\[ \langle r_{\text{FBDT}} \rangle \]

\[ \langle r_{\text{MLP}} \rangle \]
Deep-neural flavor tagger

- 10 tracks at maximum
- Sorted by momentum and grouped by charge.

**Input:** PID variables, momentum, azim. and polar angles, impact params., hit counts in trackers. Total = 140

**Training:** Pylearn2 Library ⇒ Theano ⇒ GPUs.

Effective efficiencies:

$$\varepsilon_{\text{eff}} \text{ [%]}$$

<table>
<thead>
<tr>
<th>Category</th>
<th>MC</th>
<th>Belle II*</th>
<th>Belle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category-based</td>
<td>37.2</td>
<td>34.3</td>
<td></td>
</tr>
<tr>
<td>Deep neural</td>
<td>40.7</td>
<td>34.4</td>
<td></td>
</tr>
</tbody>
</table>

* No beam bkg. and IP always at zero.
Category-based tagger with Belle data

- Splot performed with converted Belle data using $m_{bc}$ as discriminating variable.

- Full Belle 0.8 ab$^{-1}$

$$B^0 \rightarrow J/\psi K_S^0$$

$\mu = \sqrt{E_{\text{beam}}^*{}^2 - \mathbf{p}_{B^0}}$
- Belle Data distribution weighted with splot output variable (signal component).

- Nice overlap of output distribution for Belle MC and Belle data 😊.

Belle II: $\varepsilon_{\text{Eff}} = 33.6 \pm 0.5\%$ on Belle data (Assuming MC calibration).

- $\varepsilon_{\text{Eff}}(\text{Belle}) = 30.1 \pm 0.4\%$, $\varepsilon_{\text{Eff}}(\text{BaBar}) = 33.1 \pm 0.3\%$. 
Conclusions and Outlook

- Performance of category-based and of deep-neural taggers are similar with Belle MC.
- On Belle II MC without beam bkg. and beam spread, the deep-neural algorithm performs better.
- The category-based tagger has been validated on Belle data.
- For the deep-neural tagger is ongoing.
- Calibration of both taggers using Belle data (flavor-mixing measurement) is ongoing.
- Benchmark calibration with first $\sim 20\, fb^{-1}$ Belle II commissioning data possible (Preparation ongoing).
- We want to use both algorithms for better understanding of MC/Data differences.
Efficiency Calculation

- Binning ⇒ correction with real data!

- Efficiency:

  \[ \varepsilon_{\text{Eff}} = \sum_i \varepsilon_i \cdot \langle r_i \rangle^2 \]

- \( r_{\text{MC}} = 1 - 2 \cdot w_{\text{MC}} \)

- Calibration: \( r_{\text{MC}} \) linear to \( r \) Output
### Tagging variables

<table>
<thead>
<tr>
<th>Categories</th>
<th>Discriminating input variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>( \mathcal{L}_e, p^<em>, p_t^</em>, p, p_t, \cos \theta, d_0,</td>
</tr>
<tr>
<td>Int. Electron</td>
<td></td>
</tr>
<tr>
<td>Muon</td>
<td>( \mathcal{L}_\mu, p^<em>, p_t^</em>, p, p_t, \cos \theta, d_0,</td>
</tr>
<tr>
<td>Int. Muon</td>
<td></td>
</tr>
<tr>
<td>Kin. Lepton</td>
<td>( \mathcal{L}<em>e, \mathcal{L}</em>\mu, p^<em>, p_t^</em>, p, p_t, \cos \theta, d_0,</td>
</tr>
<tr>
<td>Int. Kin. Lep.</td>
<td></td>
</tr>
<tr>
<td>Kaon</td>
<td>( \mathcal{L}_K, p^<em>, p_t^</em>, p, p_t, \cos \theta, d_0,</td>
</tr>
<tr>
<td>Slow Pion</td>
<td>( \mathcal{L}_\pi, \mathcal{L}_e, \mathcal{L}_K, p^<em>, p_t^</em>, p, p_t, \cos \theta, d_0,</td>
</tr>
<tr>
<td>Fast Hadron</td>
<td>( \mathcal{L}_\pi, \mathcal{L}_e, \mathcal{L}_K, p^<em>, p_t^</em>, p, p_t, \cos \theta, d_0,</td>
</tr>
<tr>
<td>Kaon-Pion</td>
<td>( \mathcal{L}<em>K, y</em>{K\text{on}}, y_{\text{SlowPion}}, \cos \theta_{K\pi}^*, q_K \cdot q_\pi )</td>
</tr>
<tr>
<td>Maximum P*</td>
<td>( p^<em>, p_t^</em>, p, p_t, d_0,</td>
</tr>
<tr>
<td>FSC</td>
<td>( \mathcal{L}<em>{K\text{Slow}}, p</em>{\text{Slow}}^<em>, p_{\text{Fast}}^</em>, \cos \theta_T^<em>, \cos \theta_T^</em>, \cos \theta_T^<em>, \cos \theta_{\text{SlowFast}}^</em>, q_{\text{Slow}} \cdot q_{\text{Fast}} )</td>
</tr>
<tr>
<td>Lambda</td>
<td>( \mathcal{L}<em>p, \mathcal{L}</em>\pi, p_\Lambda^<em>, p_\Lambda, p_\pi^</em>, p_\pi, p_\pi, p_\pi, q_\Lambda, M_\Lambda, n_{K^0}, \cos \theta_{x_\Lambda \cdot p_\Lambda},</td>
</tr>
</tbody>
</table>

Optimized for CPU: Each Variable is calculated only once for each particle list! (108 instead of 220 calculations)

5 Particle lists: \( e, \mu, K, \pi, \Lambda \)