GPU usage plans

Giulio Dujany

IPHC - CNRS

Belle II France computing workshop

Introduction

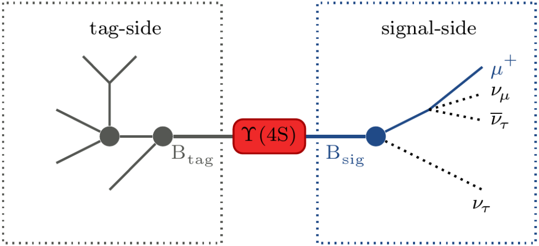
Deep learning

- Highly successful for complex problems characterised by a hierarchical representation of the data (machine vision, natural language processing, web searches ...)
- Allows good performances using "raw variables" without needing "optimal variables" defined by domain experts
- Increasing role of deep learning in particle physics

- Interest in the French Belle II groups to start exploring the usage of deep learning to tackle our problems
- Need to use GPUs to reduce the training time

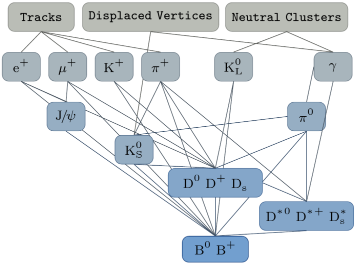
The main problem we want to tackle

- Study B decays with invisible particles in the final state
- Not possible to reconstruct the B directly its final state particles: need to reconstruct the other B_{tag} coming from the $\Upsilon(4S)$ decay to close the kinematics
- This is a complex problems characterised by a hierarchical representation of the data as the B can decay in thousands of different ways



How we want to solve it (1)

- Use DNN to fully reconstruct the decay of the B_{tag} from its final state particles without coding explicitly the possible decay modes
- A particularly promising DNN architecture we want to explore is the Graph Network [Battaglia et al., arXiv:1806.01261]
- Study benefit over algorithm based on BDT already used in Belle II



In Strasbourg an ANR project submitted on this topic plus a M2 stage in collaboration with KIT starting later this year

How we want to solve it (2)

- Reconstruct only the decays $B_{\mathrm{tag}} \to D^0 X$ exploiting the large branching fraction
- Study X with a DNN to separate B decays from background
  - presence of leptons, kaons, π^0, clusters, ...

	$\mathbf{B^+ \to}$	$\mathbf{B^0 \to}$
$\mathbf{D^0 X}$	$(8.6\pm0.7)\%$	$(8.1\pm1.5)\%$
$\mathbf{\overline{D}^0 X}$	$(79\pm4)\%$	$(47.4\pm2.8)\%$
$\mathbf{D^+ X}$	$(2.5\pm0.5)\%$	$(< 3.9\%)$
$\mathbf{D^- X}$	$(9.9\pm1.2)\%$	$(36.9\pm3.3)\%$
$\mathbf{D_s^+ X}$	$(7.9\pm1.4)\%$	$(10\pm2)\%$
$\mathbf{D_s^- X}$	$(1.10\pm0.40)\%$	$(< 2.6\%)$
$\mathbf{\Lambda_c^+ X}$	$(2\pm1)\%$	$(< 3.1\%)$
$\mathbf{\Lambda_c^- X}$	$(3\pm1)\%$	$(5.0\pm2.0)\%$

At IJC Lab a postoc and a PhD student plan to work on this

Other projects also possible

- signal/background discrimination
- muon identification
- ...

Requests for 2020

- Guesstimated 1800 hours of GPU time (accessing the batch queues)
- Plan to use GPUs only for deep learning so do not plan to code in CUDA ourself but use dedicated python library (probably Keras/Tensorflow, may have a look at pytorch if available)

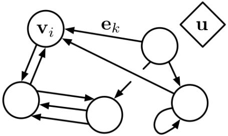
WARNING

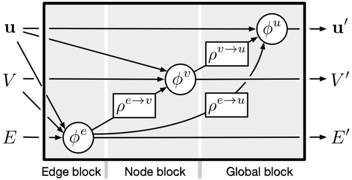
- We do not have much experience yet so for this first year we made an educated guess.
- Your feedback is more than welcome!
- This year's experience will help coming up with a more accurate estimate for next year

Thanks to Aresh Vedaee and Renaud Vernet for their help!

BACKUP

General graph network processing pipeline

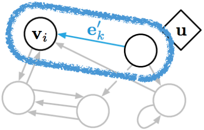




Edge block

For each edge, $\mathbf{e}_k, \mathbf{v}_{s_k}, \mathbf{v}_{r_k}, \mathbf{u}$, are passed to an "edge-wise function":

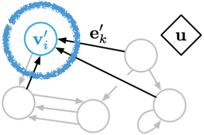
$$\mathbf{e}'_k \leftarrow \phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right)$$



Node block

For each node, $\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}$, are passed to a "node-wise function":

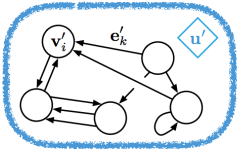
$$\mathbf{v}'_i \leftarrow \phi^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right)$$



Global block

Across the graph, $\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}$, are passed to a "global function":

$$\mathbf{u}' \leftarrow \phi^u\left(\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}\right)$$



B-tagging...

standard tagging methods: hadronic and semi-leptonic

other possibilities ? semi-inclusive, a.k.a **c-tag**...

signal side: K+ 2 leptons (especially $K\tau l$) allows looser selection in B_{tag} side ?

⇒ flavour tagging combines information as charge/momentum of lepton, charge of kaon, charge of soft pion etc... ⇒ output (qr): B or $\overline{B}$

⇒ B-tagging... but better to talk about charged B tag or neutral B tag

	$B^+\to$	$B^0\to$
$D^0 X$	$(8.6\pm0.7)\%$	$(8.1\pm1.5)\%$
$\overline{D}^0 X$	$(79\pm4)\%$	$(47.4\pm2.8)\%$
$D^+ X$	$(2.5\pm0.5)\%$	$(< 3.9\%)$
$D^- X$	$(9.9\pm1.2)\%$	$(36.9\pm3.3)\%$
$D_s^+ X$	$(7.9\pm1.4)\%$	$(10\pm2)\%$
$D_s^- X$	$(1.10\pm0.40)\%$	$(< 2.6\%)$
$\Lambda_c^+ X$	$(2\pm1)\%$	$(< 3.1\%)$
$\Lambda_c^- X$	$(3\pm1)\%$	$(5.0\pm2.0)\%$

- $B^+\to\overline{D}^0 X$ BR is high
  - reconstruct exclusively D^0 (20%~$\epsilon\times$BR, $D\to K\pi, K\pi\pi^0$, K3pi, $K_S\pi\pi$, KK...), p_{CM}...
  - exploit 'Dalitz' properties of $D\to Kn\pi$
- D^+ would be a veto !
- D_s^+: additional info (but $BR(D_s\to\phi(K^+K^-)\pi)$~2.3% only), $\phi + D^0$? ($BR(D_s\to\phi(K^+K^-)X)$~8%
- in fact think as $B^+\to\overline{D}^0[\mu\nu]X$, $\to\overline{D}^0[e\nu]X$, $\to\overline{D}^0[c\bar{s}]X$ (here $\overline{D}^0 D_s X$ or $\overline{D}^0\phi X$), $\overline{D}^0[u\bar{d}]X$
- not forgetting $B^+\to J/\psi K X$ (K^+ and K_S^0)
- study of X: presence of leptons, kaons, π^0, clusters... ⇒ BDT, DNN...

⇒ charged Btag info with a quality information, ≠ neutral Btag algorithm

13