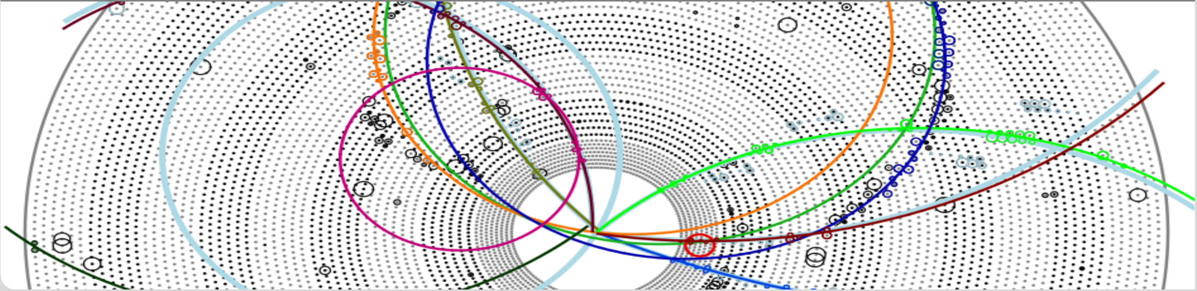
Continuum background suppression using Deep Learning for the Belle II experiment

IMPRS Recruiting Workshop

Luciana Tanzarella | 18th July 2022

MAX PLANCK INSTITUTE FOR PHYSICS - BELLE II GROUP



Politecnico di Torino (Italy), Physics of Complex Systems – Max Planck Institute for Physics

Contents

SuperKEKB

- asymmetric energy e^+e^- Super B-Factory
  $\rightarrow$ 7 GeV e^- and 4 GeV e^+
- set a new brightness world record of $(3.8 \cdot 10^{34} \mathrm{cm}^{-2} \mathrm{s}^{-1})$ in December 2021
  $\rightarrow$ high precision measurements of rare decays and CP-violation

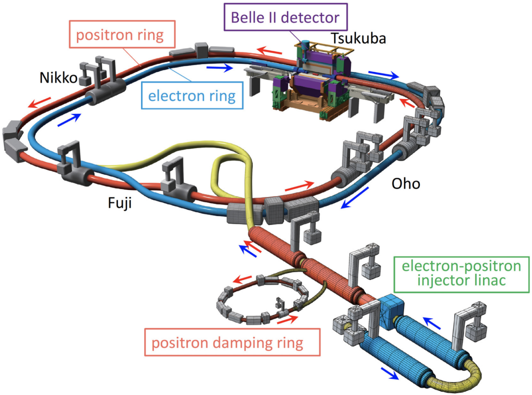


Figure: The SuperKEKB complex. From Akai, Furukawa, and Koiso n.d.

Belle II experiment

- general-purpose spectrometer for the next-generation B-factory experiment;
- made up by layered sub-components, specific to detect particles at a specific energy or trajectory.

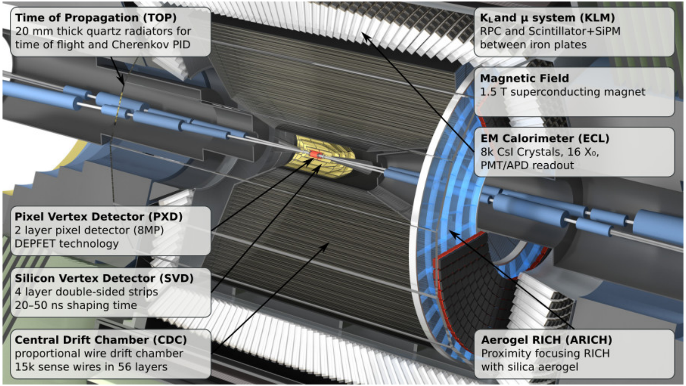


Figure: The Belle II detector. From basf2 Online Textbook, Data taking n.d.

Belle II experiment

Focus

Study charmless decays of the B meson

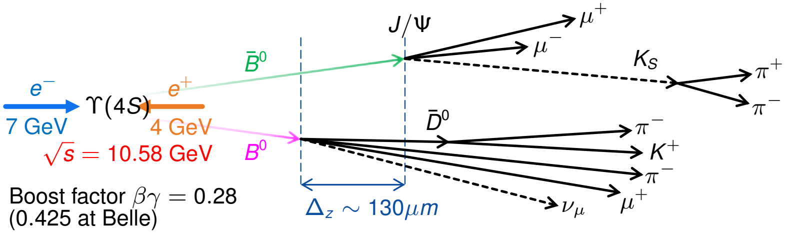
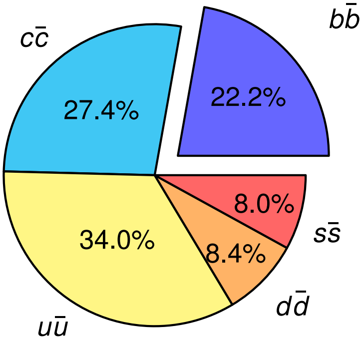


Figure: Decay at Belle II, $\Upsilon(4S)$ resonance

Continuum background

- non resonant $e^+e^- \rightarrow q\bar{q}$ events: the most common source of this combinatorial background;
- hadronisation of lighter quarks $\rightarrow u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$



Focus

$b\bar{b}$ events are relevant. All the rest is background.

Variables: topological discriminators

Due to high momentum suitable for decay to light hadrons, the continuum particles are collimated (jet-like shape).
The BB event's particles are evenly distributed (spherical shape).

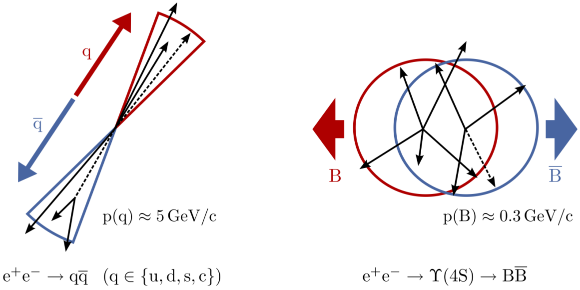


Figure: Event shapes: Continuum vs signal

$\rightarrow$ binary classification task.

Variables for Continuum Suppression

Engineered Variables (E)

- Fox Wolfram Moment
- Kakuno-Super-Fox-Wolfram variables
- Thrust
- CleoCones

Detector Level Variables (DL)

- Basic Variables (momentum, azimuthal angle and polar angle and relative uncertainties)
- Track Variables (particle ID, number of CDC hits, probability of track fit)
- Cluster Variables

Vertex Variables (V)

- distance (IP-decay vertex)

Deep Neural Networks

- Neural Networks are non-linear models for biologically inspired supervised learning
- Fundamental element: the artificial neuron (perceptron)

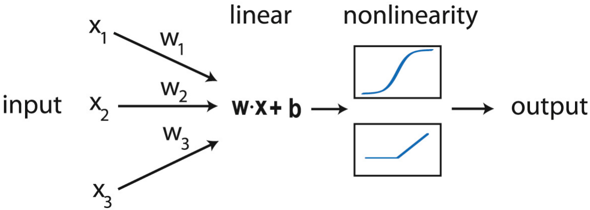


Figure: Each input node has a weighted edge that connects it to an output node. The output is the weighted average of the inputs minus a bias factor, to which an activation function is applied. Adapted from Mehta et al. 2019

$$z^{(i)} = \mathbf{w}^{(i)} \cdot \mathbf{x} + b^{(i)} \quad (1)$$

Analysis

Framework: Pytorch
Preprocessing:

- normalization of the inputs;
- turn NaN values into zeros;

Hyper-parameter tuning

Set	NHL	NL	AUC
1	1	50	0.9958
2	1	100	0.9959
3	1	300	0.9958
4	3	100	0.9957
5	3	150	0.9963
6	3	50	0.9964
7	5	50	0.9969
8	5	100	0.9969
9	6	50	0.9957
10	6	100	0.9956

Table: Hyperparameter tuning for all variables

DNN

Politecnico di Torino

- 5 sequential layers;
- ADAM optimizer;
- ReLU activation function;
- 50 nodes each layer;
- 512 events per mini-batch;
- run for 10 epochs.

Classifier	AUC
DNN(E+DL+V)	0.9969
DNN(E+DL)	0.9956
DNN(E)	0.9728

Table: AUC for each feature set

ROC curve for DNN

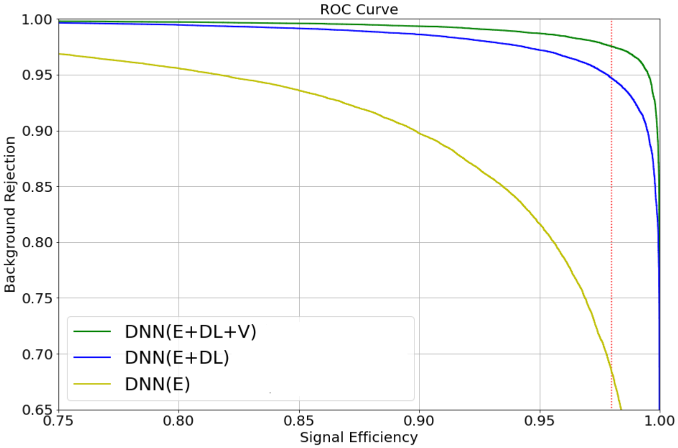
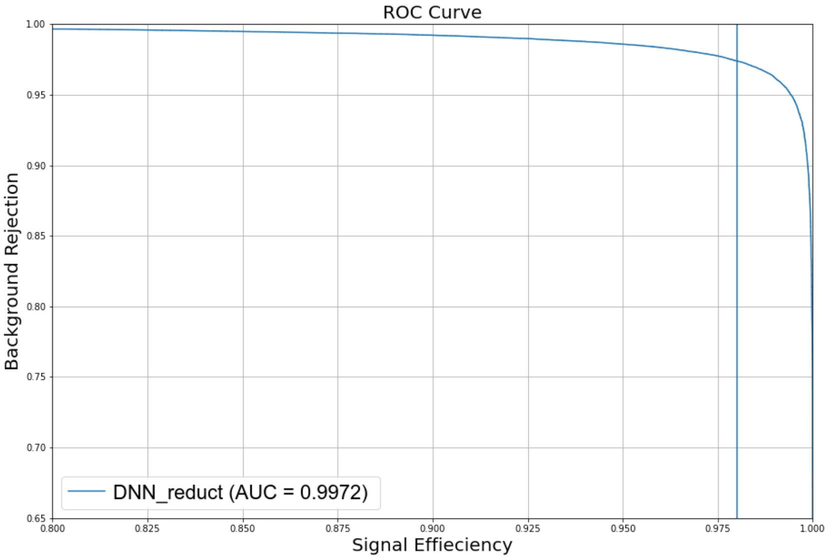


Figure: Comparison of different feature sets. The dashed vertical line represents 98% of the signal effectiveness, i.e. the minimum target to be reached.

Subleading particle approach

- Inspect the dataset and its variable sets
- starting point: 361 variables $\rightarrow$ redundant information
- By gradually reducing the variables, performance is not adversely affected:
  - engineered variables are built on low-level variables $\rightarrow$ redundancy (eliminates 61 variables)
  - IDs (identification probability) may lead to errors, as well as basic variables and cluster specific variables $\rightarrow$ 108 variables remains
  - do not eliminate the vertex variables due to their high discriminative power

ROC - Reduced features



Summary and future perspectives

- Data taken from MC14 and adapted to the task through steering;
- With a simple architecture, the technique of variable-space reduction has proved effective and promising
  $\rightarrow$ lower computational demands;
- ptimal combinations of hidden layers and neurons for each layer were obtained not to use more resources than necessary;
- Machine Learning shows great potential in pattern recognition for HEP
  $\rightarrow$ develop new architectures: Graph NN, Convolutional NN, GAN (Generative adversarial network)

Bibliography

[1] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. "SuperKEKB collider". In: Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 907 (). DOI: 10.1016/j.nima.2018.08.017.

[2] basf2 Online Textbook, Data taking. https://software.belle2.org/development/sphinx/online_book/fundamentals/02-datataking.html.

[3] Pankaj Mehta et al. "A high-bias, low-variance introduction to Machine Learning for physicists". In: Physics Reports 810 (2019). A high-bias, low-variance introduction to Machine Learning for physicists, pp. 1–124. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2019.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0370157319300766.