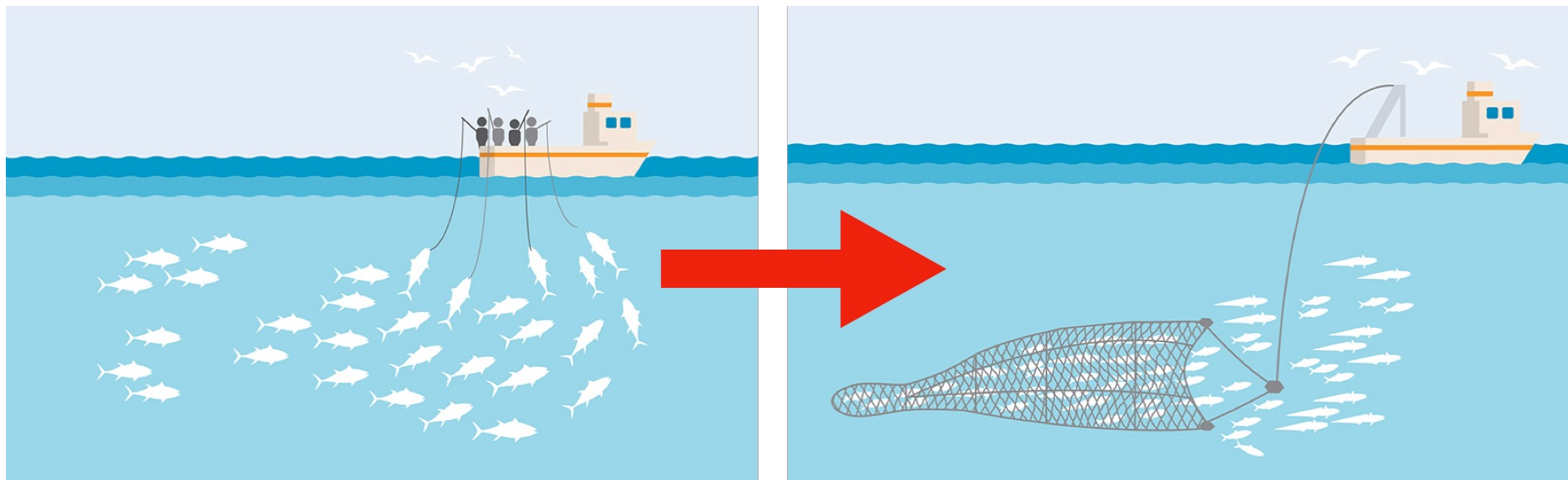


基于机器学习多分类的数据分析

李刚

所创新课题启动会

从二分类到多分类



- 数据分析是个分类问题：传统做法基本都是二分类。
- 机器学习很容易实现 多分类
 - 降低人力和时间成本
 - 利用完整数据信息
 - 有完整协方差矩阵，便于物理结果的统计诠释
 - 便于做探测器优化

希格斯事例分类

Higgs : 4 production modes , 9 decays modes

Prod\decay	cc	bb	$\mu\mu$	$\tau\tau$	$\gamma\gamma$	gg	WW	ZZ	γZ
eeH									
$\mu\mu H$									
$\tau\tau H$									
qqH									

$$\chi^2 = \sum_i \frac{(\sum \epsilon_{ij} N_j - n_i)^2}{\sigma_{n_i}^2} + \frac{(\sum_l N_l - N)^2}{\sigma_N^2}$$

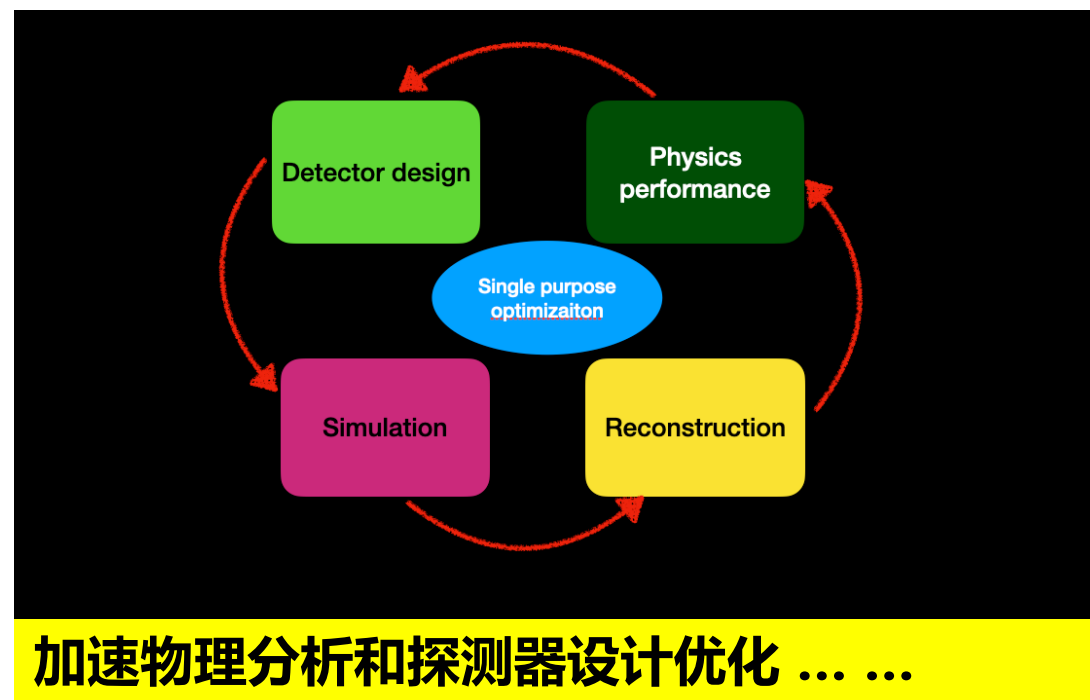
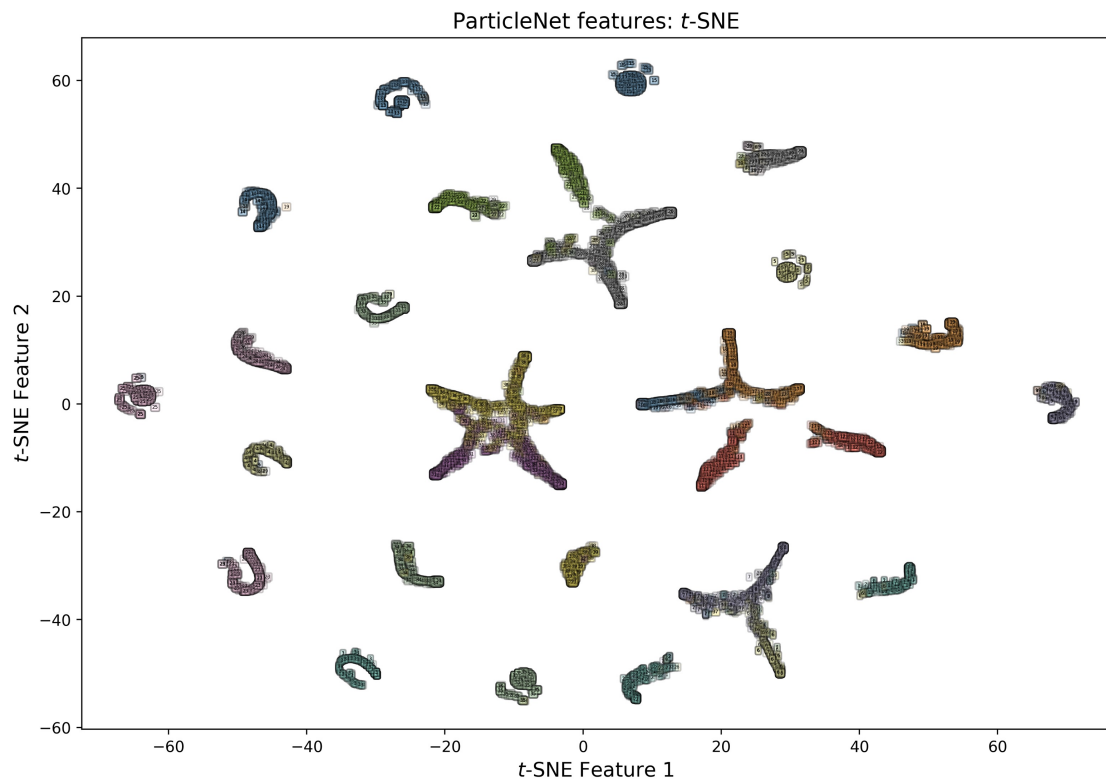
1 2 3 4 5 6 7 8 9

$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \\ n_7 \\ n_8 \\ n_9 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} & \epsilon_{14} & \epsilon_{15} & \epsilon_{16} & \epsilon_{17} & \epsilon_{18} & \epsilon_{19} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} & \epsilon_{24} & \epsilon_{25} & \epsilon_{26} & \epsilon_{27} & \epsilon_{28} & \epsilon_{29} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} & \epsilon_{34} & \epsilon_{35} & \epsilon_{36} & \epsilon_{37} & \epsilon_{38} & \epsilon_{39} \\ \epsilon_{41} & \epsilon_{42} & \epsilon_{43} & \epsilon_{44} & \epsilon_{45} & \epsilon_{46} & \epsilon_{47} & \epsilon_{48} & \epsilon_{49} \\ \epsilon_{51} & \epsilon_{52} & \epsilon_{53} & \epsilon_{54} & \epsilon_{55} & \epsilon_{56} & \epsilon_{57} & \epsilon_{58} & \epsilon_{59} \\ \epsilon_{61} & \epsilon_{62} & \epsilon_{63} & \epsilon_{64} & \epsilon_{65} & \epsilon_{66} & \epsilon_{67} & \epsilon_{68} & \epsilon_{69} \\ \epsilon_{71} & \epsilon_{72} & \epsilon_{73} & \epsilon_{74} & \epsilon_{75} & \epsilon_{76} & \epsilon_{77} & \epsilon_{78} & \epsilon_{79} \\ \epsilon_{81} & \epsilon_{82} & \epsilon_{83} & \epsilon_{84} & \epsilon_{85} & \epsilon_{86} & \epsilon_{87} & \epsilon_{88} & \epsilon_{89} \\ \epsilon_{91} & \epsilon_{92} & \epsilon_{93} & \epsilon_{94} & \epsilon_{95} & \epsilon_{96} & \epsilon_{97} & \epsilon_{98} & \epsilon_{99} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \\ N_6 \\ N_7 \\ N_8 \\ N_9 \end{pmatrix}$$

$$\sum_i B_i = 1 \quad B_i = \frac{N_i}{N}$$

CEPC 上成功实现多分类任务

图形卷积神经网络实现多分类(> 36) , 可以测量多个希格斯衰变



用于加速探测器和实验优化, 大大降低时间和人力需求, 探测器设计的快速迭代

应用场合

- $e^+e^- \rightarrow ZH$, 用 Z 来 tag Higgs
- $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi$, 用 $\pi^+\pi^-$ 来 tag J/ψ
- $e^+e^- \rightarrow \tau^+ \tau^-$, 用一个 tau 来 tag 另一个
- DD pair 用一个来 D tag 另一个
-