Probabilistic Classification of Diffractive Events

Mikael Mieskolainen

Tampere University of Technology
Helsinki Institute of Physics

CERN Summer Student Session, 17/08/2011

Diffraction in Particle Physics?

Analogy comes from optical diffraction (Landau and Pomeranchuk 1953)

Hadronic scattering processes are usually denoted either as *soft* (momentum transfer squared is "small") or *hard* (transfer is "large"), and diffractive processes are predominantly soft.

- Traditional way of classifying (and defining) diffraction is based on *Large Rapidity Gaps* (LRG).
- For hard diffraction (rare) we can use perturbative QCD, but soft diffraction is theoretically non-perturbative. For details[1]
- Classic phenomenological model for soft diffraction is the Regge Theory with *Pomeron* exchange.

[1] see e.g. V. Barone, E. Predazzi., *High-Energy Particle Diffraction*, Springer, 2002.

Elastic, Diffractive and Non-Diffractive Events

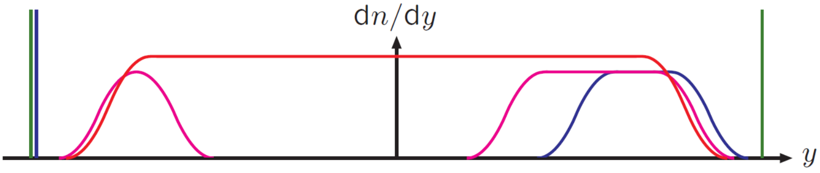

Figure: Illustration of multiplicity as a function of rapidity for different classes, [Torbjörn Sjöstrand, MCnet School, 2010]

One prediction at the LHC at CM energy $\sqrt{s} = 14$ TeV

$$\underbrace{\sigma_{tot}}_{\approx 92 \text{ mb}} \triangleq \underbrace{\sigma_{elastic}}_{\approx 20 \text{ mb}} + \underbrace{\sigma_{single-diff}}_{\approx 12 \text{ mb}} + \underbrace{\sigma_{double-diff}}_{\approx 6 \text{ mb}} + \underbrace{\sigma_{non-diff}}_{\approx 54 \text{ mb}}$$

Forward Physics at the LHC

Necessary to have forward instrumentation to study diffractive events

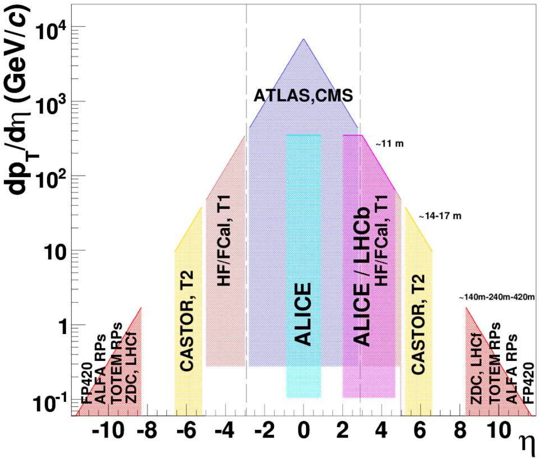

Figure: Approximate $p_t - \eta$ coverage at the LHC at $\sqrt{s} = 14$ TeV,
[David d'Enterria, Forward Physics at the LHC, 2008]

Supervised Classification

Classification of diffraction is done on an event by event. A classifier g is a mapping

$$g : \mathcal{F}^d \longrightarrow \mathcal{C},$$

where $\mathcal{F}^d$ is the d-dimensional feature space (usually Hilbert space $\mathcal{H}^d$) and $\mathcal{C}$ is the finite set of class labels.

- We want to learn this mapping, in a *supervised* manner[2], based on a training set $\mathcal{T}$ of feature vector - class label pairs:

$$\mathcal{T} = \{(\mathbf{x}_i \in \mathcal{F}^d, \mathbf{c}_i \in \mathcal{C})\}$$

[2]for introduction, see papers from Mikael Kuusela et al.

Why Probabilistic?

- The training vectors $\mathbf{x} \in \mathcal{F}^d$ contain track multiplicities, calorimeter energies, Roman Pot hits etc. of an event, over the full coverage of pseudorapidity η. These vectors are generated for different classes via Monte Carlo simulations.

Unfortunately, because of nature of diffractive events, classes are inherently overlapping in feature space!

$\Longrightarrow$ Necessary to estimate Maximum a Posteriori (MAP) probabilities $P(\text{class} = c|\mathbf{y})$ of a real event $\mathbf{y} \in \mathcal{F}^d$, to belong any of the $c \in \mathcal{C}$ classes.

This way also relative cross sections (σ_c/σ_{tot}) of different classes are estimated in higher accuracy.

Probabilistic Classification

Maximum a posteriori probabilities of a real event **y** are estimated with ℓ_1-norm regularized Multinomial Logistic Regression

$$P(\text{class} = c|\mathbf{y}) = \frac{\exp(\mathbf{w}_c^T \mathbf{y})}{\sum_{c'=1}^{|\mathcal{C}|} \exp(\mathbf{w}_{c'}^T \mathbf{y})}, \quad c = 1, \ldots, |\mathcal{C}|,$$

where weight vectors $\mathbf{w}_c$ are trained with the set $\mathcal{T}$ using convex optimization techniques. Algorithm can be also *kernelized* to achieve non-linear decision boundaries.

- Reqularization in training phase with ℓ_1-norm induces *sparsity* into $\mathbf{w}_c \Rightarrow$ algorithmic feature (detector) selection.

Conclusions

Currently working with the classification of CDF (Tevatron) zero-bias data, interesting to see kinematic distributions and relative cross sections as a result of classifications.

- At the LHC, especially the TOTEM experiment is involved in diffractive physics (forward physics), but also all the big experiments (CMS, ATLAS, ALICE...).
- Supervised classification depends on Monte Carlo models. To be independent of MC, one needs unsupervised machine learning.
- Diffractive physics can be seen as a demanding and interesting test bench of theoretical models.