



PROGRAMME
DE RECHERCHE
NUMÉRIQUE
POUR L'EXASCALE

WP3 - ML-based data analytics

Thomas Moreau & Bruno Raffin

Identified applications

From discussion with Gysela / SKA / Coddex

Event detection and tracking

- Finding and tracking patterns
- Finding change points

Anomaly detection

- Modeling nominal data
- Finding model deviation

Data compression

- For storage/comm'
- For anomaly detection

Challenges

- *Data cannot be stored* -> need learning algorithms that can handle streams of data
- *Data is distributed* -> need models that work on subdomains
- *Labelling is costly* -> unsupervised learning/transfer learning
- *In situ* -> need to be fast enough and have limited auxiliary memory

Event detection

Codex

- Hot spot
- only few event per simulation/ few simulations
- Used for steering

Tokam2D/Gysela

- Burst of density
- many events per frame
- Trajectory are of interest
- Used for steering

SKA

- Fast radio bursts
- few events but many "frames"
- The trajectory of events is of interest

In Mesh data

In Nd-array that evolve through time

Distributed data

Single node data

Roadmap: adapt Computer Vision literature to physical signals

Data-driven compression

- Necessary to do compression to store/communicate the simulation result (big Nd-array)
- But compression can be adapted to specifically compute some diagnostic (statistics)

This is the interest of data-driven compression

Gysela/Tokam2d:

- Compression of the 3D information to have the best reconstruction?
- The compression model needs to run with the distributed data

Machine Learning Motifs

From a ML perspective

Learning from distributed data

Learn a model that makes local decision based on Nd-array data partitioned into sub-domains, by minimizing communication and auxiliary memory consumption.

Unsupervised event-tracking

In large Nd-array evolving in time, some patterns are repeating (spatially) and moving (time). We would like to identify them and track them automatically, if possible with low memory/latency.

Anomaly detection

Detect deviation of the simulation with normal behavior to be able to stop simulation before numerical instability.



PROGRAMME
DE RECHERCHE

NUMÉRIQUE
POUR L'EXASCALE

Retrouvez toutes nos actualités

 NumPEX