Perspectives on High-Performance Computing in a Big Data World – Part D

The 28th International Symposium on High-Performance Parallel and Distributed Computing

Geoffrey Fox | June 27, 2019

gcf@indiana.edu, http://www.dsc.soic.indiana.edu/, http://spidal.org/
Learning Model Details
Agents and Time-Series Case Studies
a) **Learning Agent Behavior** One has a model such as a set of cells as agents modeling a virtual tissue. One can use ML to learn dynamics of cells replacing detailed computations by ML surrogates.

- As can be millions to billions of such agents the performance gain can be huge as each agent uses same learned model..
- This is MLaroundHPC for cells but MLAutotuning for multi-cell (tissue) phase.
b) Learning Effective Potentials or Interaction Graphs

An effective potential is an analytic, quasi-empirical or quasi-phenomenological potential that combines multiple, perhaps opposing, effects into a single potential.

Learning Model Details: Learning Effective Potentials or Interaction Graphs

• This is classic coarse graining strategy
• Deep Learning replacing dimension reduction techniques
Take the case where we have “videos” recording observational data i.e. data is a high dimensional (spatial extent) time series

(c) Learning Agent Behavior – a Predictor-Corrector approach Here one time steps models and at each step optimize the parameters to minimize divergence between simulation and ground truth data.

Example: produce a generic model organism such as an embryo. Take this generic model as a template and learn the different adjustments for particular individual organisms.

• Build on Ride hailing work
• Current state of the art expresses spatial structure as a convolutional neural net and time dependence as recurrent neural net (LSTM)
Yan Liu@USC ICML 2019 Time Series workshop

- Beijing:
  - 1296 regions, 19M samples
  - 10 months in 2017

- Shanghai
  - 896 regions, 13M samples
  - 10 months in 2017

Both spatial and temporal correlations modeling are necessary
- Removing either graph component leads to significantly worse performance.
- With CGRNN, ST-MGCN achieves the best performance.
Workshop on Streaming Systems and Realtime Machine Learning (STREAM-ML)

Dec 9-12, 2019 · Los Angeles, California

As a successor to the STREAM 2015 & 2016 workshops, STREAM-ML aims to advance decision-making and control over complex systems by applying machine learning techniques to streams of real-time data.

In conjunction with
IEEE International Conference on Big Data 2019

Building upon prior workshops, with support from
Actually Really Need To do Everything Simultaneously in MLaroundHPC
Simulating Biological Organisms (with James Glazier @IU)

Learning Agent Behavior
Replace components by learned surrogates
(Reaction Kinetics Coupled ODE’s)

All steps use MLAutotuning

Smart Ensembles

Predictor-Corrector
**MLControl**

a) **Experiment Control** Using simulations (possibly with HPC) in control of experiments and in objective driven computational campaigns. Here the simulation surrogates are very valuable to allow real-time predictions. Applied in Material Science and Fusion

b) **Experiment Design** One of the biggest challenges of models is the uncertainty in the precise model structures and parameters. Model-based design of experiments (MBDOE) assists in the planning of highly effective and efficient experiments – it capitalizes on the uncertainty in the models to investigate how to perturb the real system to maximize the information obtained from experiments. MBDOE with new ML assistance identifies the optimal conditions for stimuli and measurements that yield the most information about the system given practical limitations on realistic experiments.