Project DisCo: Physics-based discovery of coherent structures in spatiotemporal systems

Adam Rupe\(^1\)\(^2\), Karthik Kashinath\(^2\), Nalini Kumar\(^3\), James P. Crutchfield\(^1\), Ryan G. James\(^1\), and Prabhat\(^2\)

\(^1\)Department of Physics
Complexity Sciences Center
University of California, Davis

\(^2\)NERSC
Lawrence Berkeley National Laboratory

\(^3\)Intel®

Big Data Summit 2018
Project DisCo

- The Science Problem
- The Theory
- The Computation
- The Unsupervised Ladder
- The HPC Challenge

John Travolta in 'Saturday Night Fever' (Paramount)
Project DisCo

- The Science Problem
- The Theory
- The Computation
- The Unsupervised Ladder
- The HPC Challenge

on track for SC’19 Gordon Bell submission
**The Science Problem**

Unsupervised detection (segmentation) of spatiotemporal structures in climate
The Science Problem

Unsupervised detection (segmentation) of spatiotemporal structures in climate

[Image of a map with highlighted areas, possibly showing climate patterns]
The Theory: Computational Mechanics

Think of this as a *physics-based* machine learning technique

Segmentation achieved through use of *local causal states*
The Theory: Computational Mechanics

Think of this as a \textit{physics-based} machine learning technique

Segmentation achieved through use of \textit{local causal states}

Causal states defined through the \textit{causal equivalence relation}

\[ \text{past}_i \sim \epsilon \ \text{past}_j \iff \Pr(\text{Future}|\text{past}_i) = \Pr(\text{Future}|\text{past}_j) \]

\textit{Lightcones} used as local notions of past and future
The Computation

Goal - reconstruction of *causal equivalence relation* from data

\[ \ell_i^- \sim_\epsilon \ell_j^- \iff \Pr(L^+|\ell_i^-) = \Pr(L^+|\ell_j^-) \]
The Computation

Goal - reconstruction of causal equivalence relation from data

$$\ell_i^- \sim_\epsilon \ell_j^- \iff \Pr (L^+ | \ell_i^-) = \Pr (L^+ | \ell_j^-)$$

- Training
  1. Reconstruct morph($\ell_i^-$) = Pr ($L^+ | \ell_i^-$)

    ▶ extract ($\ell^-$, $L^+$) pairs from sample fields
    ▶ for real-valued fields, need to cluster space of lightcones
The Computation

Goal - reconstruction of causal equivalence relation from data

\[
\ell_i^- \sim_\epsilon \ell_j^- \iff \Pr(L^+|\ell_i^-) = \Pr(L^+|\ell_j^-)
\]

▷ Training

1. Reconstruct \(\text{morph}(\ell_i^-) = \Pr(L^+|\ell_i^-)\)
   - extract \((\ell^-, \ell^+)\) pairs from sample fields
   - for real-valued fields, need to cluster space of lightcones

2. Cluster together pasts with same morph - resulting clusters are local causal states
   - gives \(\epsilon\)-map; \(\epsilon(\ell_i^-) = \xi_{\ell_i^-} = \{\ell_j^- : \ell_j^- \sim_\epsilon \ell_i^-\}\)
The Computation

Goal - reconstruction of causal equivalence relation from data

\[ \ell_i^- \sim_\epsilon \ell_j^- \iff \Pr(L^+|\ell_i^-) = \Pr(L^+|\ell_j^-) \]

- Training
  1. Reconstruct \( \text{morph}(\ell_i^-) = \Pr(L^+|\ell_i^-) \)
     - extract \((\ell^-, \ell^+)\) pairs from sample fields
     - for real-valued fields, need to cluster space of lightcones
  2. Cluster together pasts with same morph - resulting clusters are local causal states
     - gives \( \epsilon \)-map; \( \epsilon(\ell_i^-) = \xi_{\ell_i^-} = \{\ell_j^- : \ell_j^- \sim_\epsilon \ell_i^-\} \)

- Inference
  1. Use \( \epsilon \)-map to perform causal filtering; \( \mathbf{x} \rightarrow S = \epsilon(\mathbf{x}) \)
     - Segmentation semantics from structural properties of \( S \)
The Unsupervised Ladder

No labeled data – no error metric to optimize
The Unsupervised Ladder

No labeled data – no error metric to optimize

Built from physical theory

- validation established using physical principles

broken symmetry
Step 1 – Cellular Automata

**Step 2 – Coupled Map Lattices**
Step 3 – Vortex Shedding

K = 3

K = 11
**Step 4 – Bickley Jet**

Candidate data set for SC’19 GB submission

Step 5 – Transitional / Turbulent Flow

Candidate data set for SC’19 GB submission
The HPC Challenge

- Clustering very high-dimensional lightcone data
  - kmeans vs dbscan
- Multi-node clustering in Python
- Efficient use of memory
OUTCOMES

Completed Papers:


- Accepted for publication in Chaos: An Interdisciplinary Journal of Nonlinear Science


Planned Manuscripts:

▶ Structural Semantics of Local Causal States, Part 1: Contamination.
▶ Structural Semantics of Local Causal States, Part 2: Coherence Detection.
▶ Local Causal States and Lagrangian Coherent Structures.
Outcomes

Presentations:

- Seminar – Center for Nonlinear Dynamics, UT Austin, 2016
- Talk – APS Far West Section, 2016
- Poster – AGU Fall Meeting, 2016
- Talk – Dynamics Days, 2017
- Talk – 7th Annual UC Davis Math Conference, 2017
- Poster – 7th International Workshop on Climate Informatics, NCAR, 2017
- Poster – Intel HPC Developer Conference, 2017
- Talk – 70th Annual Meeting of the APS Division of Fluid Dynamics, 2018
- Poster – Dynamics Days, 2018
- Poster – Intel AI Dev
- Talk – Information Engines at the Frontiers of Nanoscale Thermodynamics, Telluride Science Research Center
Thank You!