Challenges for AI in Computational Sustainability

Carla P. Gomes
Cornell University

Accelerating Science: A Grand Challenge for AI
AAAI Fall Symposium
November 17-19, 2016
Arlington, VA, United States
To nucleate the Computational Sustainability field

To identify a number of core research directions for maximal impact, both in terms of Computer Science and Sustainability.

1st Expeditions in Computational Sustainability (2008)

2nd Expeditions: Large-Scale Research Network (2015) for Expanding the Horizons of Computational Sustainability

Thank you!

CompSustNet

125+ faculty, students, and collaborators!!!
12 US Academic Institutions
10 US Organizations (Gov and NGOs)
and 16 International Organizations (Academic Institutions, Gov, and NGOs)
Computational Sustainability: Advancing Computer Science and Sustainability Science

Computational Sustainability is a new interdisciplinary field that aims to advance computational methods to help balance economic, environmental, and societal needs for sustainable development.
Computational Sustainability: Advancing Computer Science and Sustainability Science

Computational Sustainability is a new interdisciplinary field that aims to advance computational methods to help balance economic, environmental, and societal needs for sustainable development.

1. New challenging problems
2. New formalisms and concepts from other disciplines

→ New Core Paradigmatic problems in Comp. Sci.

→ Broadening the Scope & Diversity in CS

→ Societal Impact

Computational Thinking that will provide new insights, methodologies, and solutions to sustainability problems.
Challenges in Decision-Making for Managing Complex Natural and Man-Made Eco-Systems

**Economic Model**
- e.g. combinatorial auctions
- Wind, solar, water
- Plants and animals
- Ev and grid batteries

**Non-linear Dynamics**
- Nonlinear Dynamics: Logistics map (simple population model)
  \[ x_{t+1} = r x_t (1 - x_t), \]
  \( r \) is the growth rate
  For certain values of \( r \) → chaotic behavior
  (difficult to make predictions even for deterministic models)

**Decision-Making (Management and Prediction)**

**Resources**
Challenges in Decision-Making for Managing Complex Natural and Man-Made Eco-Systems

Non-linear Dynamics

Logistics map (simple population model)
\[ x_{t+1} = r x_t (1 - x_t), \]
\( r \) is the growth rate
For certain values of \( r \rightarrow \) chaotic behavior
(difficult to make predictions even for deterministic models)

Uncertainty
Gaussian noise, heavy-tails, bursts, rare-events, lagged processes, temporal and model uncertainty, ...

Resources
Plants and animals
Wind, solar, water
Ev and grid batteries

Decision-Making (Management and Prediction)

Economic Model

Complex Natural and Man-Made Ecosystem
Challenges in Decision-Making for Managing Complex Natural and Man-Made Eco-Systems

Economic Model

Complex Natural and Man-Made Ecosystem

Decision-Making (Management and Prediction)

Non-linear Dynamics

Multiple temporal and spatial scales

- Multiple climate scales
- Multiple ecological scales
- Multiple energy scales...

Resources

- Wind, solar, water
- Plants and animals
- Ev and grid batteries

Uncertainty

Nonlinear Dynamics: Logistics map (simple population model)

\[ x_{t+1} = r x_t (1 - x_t), \]

r is the growth rate

For certain values of r →
chaotic behavior
(difficult to make predictions even for deterministic models)

Gaussian noise, heavy-tails, bursts, rare-events, lagged processes, temporal and model uncertainty, ...

Sustainable development

Multiple time scales

Meeting present needs (short-term) without compromising the needs of future generations (long-term)
Challenges in Decision-Making for Managing Complex Natural and Man-Made Eco-Systems

Economic Model

Decision-Making (Management and Prediction)

Non-linear Dynamics

Multiple Agents Often with Conflicting Preferences

Uncertainty

Nonlinear Dynamics:
Logistics map
(simple population model)

\[ x_{t+1} = r x_t (1 - x_t) \]

\( r \) is the growth rate
For certain values of \( r \) → chaotic behavior
(difficult to make predictions even for deterministic models)

Gaussian noise, heavy-tails, bursts, rare-events, lagged processes, temporal and model uncertainty, ...

Multiple Agents

Complex Natural and Man-Made Ecosystem

Multiple temporal and spatial scales

Ev and grid batteries

Plants and animals

Wind, solar, water

Resources

 Sustainable development
Multiple time scales
Meeting present needs (short-term) without compromising the needs of future generations (long-term)

Uncharted Territory for Computer Scientists

Optimization problems
(discrete and continuous variables) with underlying complex dynamics under uncertainty over multiple scales and multiple objectives
CompSustNet: 3 Core Sustainability Themes
Over 20 Novel Interdisciplinary Research Projects (IRPs)

I Conservation and Biodiversity

- Bird Conservation
- Prioritizing Conservation Decisions
- Wildlife Corridors
- Protecting Endangered Species

II Balancing Socio-economic Needs

- Improving Weather Observations in Africa
- Protecting Migratory Herders in Africa
- Social-economic ecological corridor (Ecuador)
- Poverty Mapping

III Renewable Energy, Accelerating Discovery of Sustainable materials, and Smart Cities

- Accelerating the Discovery of New Sustainable Energy Materials
- Energy Storage and Integration of Renewables In Smart Grid
- Smart Bike Sharing

Wide range of sustainability applications:
To build a coherent and cohesive computer science sub-field it’s paramount to go deep in terms of Cross-Cutting Core Computational Problems and apply them to different sustainability areas.
Accelerating Materials Discovery for Sustainable Energy

Artificial Tree
(solar-fuel generator)
White House Materials Genome Initiative

**Goal:**
Accelerate the pace and reduce the cost of discovery and deployment of advanced material systems.

Very exciting new research area for Computer Science with tremendous potential for impact.
Accelerating the Development of Solar Fuel Generators

“... existing energy approaches [...] will not be enough to secure our energy future. Meeting the challenge will require [...] scientific breakthroughs in new materials....”


Our goal
accelerate the development of solar fuel generators

Solar fuel Cell
Artificial Tree
(solar-fuel generator)

Solar fuels can be substantially produced and stored
Discovering better catalysts for solar fuel generators

Co-sputtering
(similar to atomic spray painting)

How to determine the crystal structure of the materials, based on the X-ray diffraction patterns?

Simultaneous synthesis of thousands of materials

Silicon Wafer

FCC Crystal Structure

Goal:
Achieve High-Throughput Crystal Structure Identification

Difficult: Often X-ray diffraction patterns correspond to a mixture of crystal structures

Challenging to un-mix the X-ray diffraction patterns

Crystal Structure Map
Problem:
Infer the crystal structures of the materials from the X-ray diffraction patterns (currently a largely manual process, requiring expert knowledge)

Source Separation Problem

High-Throughput Materials Discovery

Rapid characterization of thousands of materials

(38% M1, 45% M2, 17% M3)

Cornell High-Energy Synchrotron

X-ray Diffraction

Beam intensity

Diffraction angle
Crystal Structure Map Problem

Standard techniques fail to capture the underlying physics of the phenomena.

Rich set of combinatorial constraints to capture the physics and the relationships among thousands of X-ray diffraction patterns.

Computational Synthesis:
Integration of machine learning techniques with constraint and probabilistic reasoning, sampling, and optimization techniques.
Crystal Structure Map Problem

Standard techniques fail to capture the underlying physics of the phenomena

Rich set of combinatorial constraints to capture the physics and the relationships among thousands of x-ray diffraction patterns.

Computational Synthesis:
Integration of machine learning techniques with constraint and probabilistic reasoning, sampling, and optimization techniques

Human Computation - Crowdsourcing, minimal visual pattern information, based on the heatmaps of the X-ray diffraction patterns, can lead to further speed ups (from hours to minutes)

Beautiful but challenging research problem for CS! Solving it will accelerate dramatically the discovery of new materials for solar fuels!
High-Throughput Crystal Structure Discovery

Planning and Design of Experiments: Robotics Lab

Active Learning

Data Mining and Machine Learning

Sat Modulo Theory Formal Specification

Probabilistic, Constraint Reasoning & Optimization, Matrix Factorization

Underlying Physics (e.g., XRD Shifts, known phases, lattice structure and lattice parameters, etc)

Human Input: Intermediate Level

Human Computation Expert and Non-Expert

Scientific Literature & Databases

Human Input General Public, K-12

Theory

http://www.udiscover.it

Planning and Design of Experiments:

Active Learning

Probabilistic, Constraint Reasoning & Optimization, Matrix Factorization

Underlying Physics (e.g., XRD Shifts, known phases, lattice structure and lattice parameters, etc)
Decomposition:
Slices of Sample Points
Different Visualizations of X-Ray Diffraction Patterns

Heatmap of Samples a-f

Sample a
Heatmap of the XRD pattern
Spectrogram of the XRD pattern
Detected peaks

Element 3
Element 2
Element 1

(60% Element 1, 20% Element 2, 20% Element 3)
Human Computation Framework for Boosting Combinatorial Solvers

Top: Human-Machine Framework

Decompose → Crowdsourc → Aggregate → Select → Solve → Evaluate

Edge matching puzzle

Input Data → Sub-problems: Human-Intelligent Tasks (HITs) → Completed HITs → HIT-Based Hard/Soft Constraints → Solver Solution → Contradiction Time-Out → Accept/Reject/Repeat

Bottom: Human Intelligent Tasks (Hits)

1.a) 1.b) 2.b) 2.b)
Human input can dramatically speed up the performance of combinatorial optimization methods (identification of backdoor variables).

Approaches leverages the complementary strength of human input, providing global insights into problem structure, and the power of combinatorial solvers to exploit complex constraints.

Le Bras, Bernstein, Gomes, Selman and van Dover 2013
Real System: Pd-Ta-Rh system

198 synchrotron x-ray diffraction patterns
Task #116 (1/17)

Research Participant Disclosure: This task is part of a study being conducted by Cornell University researchers and data will be collected for research purposes.

Identify Patterns of Vertical Lines

Identify the most prominent pattern (Pattern A) and, if applicable, up to two additional patterns (Pattern B and Pattern C) intersecting the Target Row. Mark the most definite and identifiable vertical lines/blobs by clicking on them with the appropriate mode selected.

桀 Instructions (click to show/hide)
桀 Example (click to show/hide)

Definitely in Pattern A  Definitely in Pattern B  Definitely in Pattern C

Messages:

Emphasize Strong Lines  Balanced Colors  Emphasize Faint Lines  Hover to Hide Markers

Any comments or notes on this particular image?

Any feedback on the instructions or the design of the task?
84% participants answered that this slice involves 8 or more peaks
Incorporating Insights form the Crowd into AgileFD, a Convolutional Matrix Factorization Algorithm

Our approach with Human Input

Normalized phase fraction using scattering intensities for the corresponding ICSD phases and AgileFD shift parameter
Pd-Rh-Ta System: Discovery of a New Material

Pd-Rh-Ta System: Discovery of a New Material

Pd$_{0.14}$Rh$_{0.40}$Ta$_{0.46}$ \(\rightarrow\) best known Pt-free catalyst for methanol oxidation
CompSustNet’s Cross-Cutting Computational Themes
Cross-Cutting Computational Theme: Pattern Decomposition in Big Data

Our Philosophy:
Students should work on similar computational problems in different domains.
Grad students working on materials discovery also work on the elephant calls.

Similar computational problems:
Both involve the decomposition of a signal (X-rays / audio) into patterns (crystal structures / types of sounds)

Elephant Listening Project
Project to monitor elephant populations by analyzing forest recordings

We represent these cross-cutting computational themes with colored “subway lines.”
Pattern Decomposition in Big Data is the blue line.
Subway Lines: Examples of Cross-Cutting Computational themes and Interactions of some of CompSustNet’s projects

- Pattern Decomposition in Big Data
- Citizen Science/Crowdsourcing
- Agents: Mechanism Design
- Large Scale Spatio-Temporal Modeling and Prediction
- Stochastic, Probabilistic Inference, and Optimization
- Large Scale Sequential Decision Making

Wind & Solar Forecasting

Characterization of Power Grid State

Electricity Usage Disaggregation

Elephant Call Detection

Flight Call Detection

Low-Cost Weather Stations in Africa (TAHMO)

High Throughput Materials Structure Identification

High-throughput Plant Phenotyping

New Forest Cicada Detection

Landscape-Scale Conservation and Rural Communities (Ecuador)

Design and Control of Electricity Storage Systems

Poverty Mapping and Targeting

Large-scale Socio-Economic Modeling of Pastoralists’ Movements (Kenya)

Green Security Games: Preventing Poaching and Illegal Fishing

Economic Dispatch of Power Generation: High Renewables Penetration

National Wildlife Refuge System

Population Estimation: Spatial Capture-Recapture

Bike Sharing

Citizen Science

Avicaching, Estimating Bird Populations and Migrations

Monitoring Sea Star Wasting Disease Outbreak

Economic Dispatch of Power Generation: High Renewables Penetration

Protection Assistant for Wildlife Security

Dynamic Precision Bird Conservation

Low-Cost Weather Stations in Africa (TAHMO)

www.Udiscover.lt
Sustainability Themes:
Biodiversity and Conservation
Balancing Socio-Economic Needs

Computational Themes:
Small Data / Big Data:
Citizen Science, Learning, Mechanism Design and Dynamical Stochastic Optimization
Citizen Science plays a key role in Bird Conservation

300K+ volunteer birders
300M+ bird observations
22M+ hours of field work (2500+years)

Distribution of eBird Observations in the US

Challenge:
Biased data, mainly from urban areas
Prevalent problem in citizen science:
Collected data are often aligned with the participants’ preferences rather than scientific objectives.

How to incentivize Citizens to visit under-sampled areas?

Very Successful in Two US Counties (19% shift to undersampled areas in a 6 month period)

Incentives:
- Avicaching points,
- leaderboards
- Lotteries (e.g. binoculars.)

Yexiang Xue, Ian Davies, Daniel Fink, Christopher Wood, Carla P. Gomes· AAMAS 2015, CP 2016, NIPS 2016
Bird Distributions
Machine Learning and Citizen Science

State of the Birds Report
(officially released by Secretary of Interior)

Novel Approaches
To Conservation
Based on eBird Models

Distribution Models for
400+ species with weekly estimates
at fine spatial resolution (3km²)

Bird Observations

<table>
<thead>
<tr>
<th>Land Cover</th>
<th>Weather</th>
<th>Remote Sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>300K+ volunteer birders</td>
<td>300M+ bird observations</td>
<td>22M+ hours of field work (2500+years)</td>
</tr>
</tbody>
</table>

Environmental Data

- 80,000+ CPU Hours (~ 10 Years!!!)

Adaptive Spatio-Temporal Machine Learning Models and Algorithms (STEM Models)

- Relate environmental predictors to observed patterns of occurrences and absences

Patterns of occurrence of the Northern Pintail for different months of the year
Source: Daniel Fink

Bird Distribution Models, Revealing, at a fine resolution, Species’ Habitat Preferences
Farmers submit bids to TNC to keep the rice fields flooded during short periods of bird migration in California.

Radically novel, dynamic, way of doing bird conservation. Only possible because of advanced computational methods.
Poverty Mitigation in Africa: Big Data for Africa

Improving Weather, Rangeland, and Forage Maps in Africa
to protect farmers and herders

Africa is very poorly sensed
(limited environmental data, vegetation maps, only a few reliable weather stations)

Herders Submit Vegetation Images and Surveys with Smartphones:
incentives: real money (small for us, good money for pastoralists)

3 month Pilot project:
→ 100,000+ surveys
Cross-Cutting Computational Theme

Findings can be transferred across domains!

Mechanism Design:
Principal-Agent Game-Theory Model

Landscape Connectivity for Andean Bears (Ecuador)

How to incentivize Citizens to visit under-sampled areas?

Computationally, these problems have similarities.

We represent these cross-cutting computational themes with colored “subway lines.”

Mechanism design is the baby blue line.

Where to place patrols to prevent poaching and illegal fishing?
Subway Lines: Examples of Cross-Cutting Computational themes and Interactions of some of CompSustNet’s projects

- Pattern Decomposition in Big Data
- Citizen Science/Crowdsourcing
- Agents: Mechanism Design
- Large Scale Spatio-Temporal Modeling and Prediction
- Stochastic, Probabilistic Inference, and Optimization
- Large Scale Sequential Decision Making

- Nitrogen Management
- Adapt-N
- TAHMO

- Poverty Mapping and Targeting
- Economic Dispatch of Power Generation: High Renewables Penetration
- Large-scale Socio-Economic Modeling of Pastoralists’ Movements (Kenya)
- Bike Sharing
- National Wildlife Refuge System
- Population Estimation: Spatial Capture-Recapture

- Wind & Solar Forecasting
- Characterization of Power Grid State

- Electricity Usage Disaggregation
- High Throughput Materials Structure Identification
- Improved Rangeland Mapping in East Africa, Citizen Science

- Flight Call Detection
- High-throughput Phenotyping

- Low-Cost Weather Stations in Africa (TAHMO)

- Principal-Agent Game-Theory Model
- Landscape-Scale Conservation and Rural Communities (Ecuador)

- New Forest Cicada Detection

- Monitoring Sea Star Wasting Disease Outbreak

- eBirders
- GrazeIt

- Dynamic Precision Bird Conservation
- Citizen Science/Avicaching, Estimating Bird Populations and Migrations

- eButterfly
- eBirder

- www.Udiscover.lt

- Offline: Tropo-Birds, collect info
CompSustNet: Large scale research network for Expanding the Horizons of Computational Sustainability

THANK YOU 😊!

125+ faculty, students, and collaborators!!!