Machine Learning and Decision Making for Sustainability

Stefano Ermon
Department of Computer Science
Stanford University

April 12
Overview

Stanford Artificial Intelligence Lab

Fellow, Woods Institute for the Environment

Big Data

Technology

Push

Computational Sustainability

Society

Pull

Sensing revolution

Artificial Intelligence
Vision: sustainability challenges as control problems

Algorithmic challenges and opportunities at every step

- Data acquisition and interpretation
- Model fitting
- Decision making and policy optimization
Computational Sustainability

Poverty traps

Decision making and optimization

natural resources management

Water and weather systems modeling

Large unstructured datasets

Machine Learning

Materials discovery for energy applications
Summary

• Introduction
• Machine Learning for Public Policy
• AI for Sustainable Energy
• Conclusion
UN’s Global Goals for Sustainable Development

The 2030 Development Agenda (*Transforming our world*)

1. End extreme poverty
2. Fight inequality & injustice
3. Fix climate change
Data scarcity

- Expensive to conduct surveys
- Poor spatial and temporal resolution
- Questionable data quality
Satellite imagery is low-cost and globally available

Simultaneously becoming cheaper and higher resolution
(DigitalGlobe, Planet Labs, Skybox, etc.)
What if... we could \textbf{infer} socioeconomic indicators from large-scale, remotely-sensed data?
Standard supervised learning won’t work

- **Lots** of unlabeled data (images)
- Very little **labeled training data** (few thousand data points)
- Nontrivial for humans (hard to crowdsource labels)
Transfer learning: Use knowledge gained from one task to solve a different (but related) task
Nighttime lights as proxy for economic development
Step 1: Predict nighttime light intensities

A. Satellite images

B. Nighttime light intensities

Deep learning model

C. Poverty measures
Training data on the proxy task is plentiful

Labeled input/output training pairs

( )  
Low nightlight intensity

...  

( )  
High nightlight intensity

Millions of training images

training images sampled from these locations
Images summarized as low-dimensional feature vectors

**Inputs:** daytime satellite images

**Outputs:** Nighttime light intensities

Convolutional Neural Network (CNN)

\[ f_1, f_2, \ldots, f_{4096} \]
Model learns relevant features automatically

No supervision beyond nighttime lights - no labeled example of what a road looks like was provided!
Transfer Learning

**Inputs:** daytime satellite images

**Feature Learning**

**Outputs:** Nighttime light intensities

{Low, Medium, High}

**Target task**

Socioeconomic outcomes

- Nonlinear mapping

- $f_1$
- $f_2$
- \ldots
- $f_{4096}$
We can differentiate different levels of poverty

2 indicators:
• Consumption expenditures
• Household assets

We outperform recent methods based on mobile call record data

Blumenstock et al. (2015) Predicting Poverty and Wealth from Mobile Phone Metadata, Science
Models travel well across borders

Models trained in one country perform well in other countries

Can make predictions in countries where no training data exists
Scalable High Resolution Poverty Maps

Run the model on about 500,000 images from Uganda:

Scalable and inexpensive approach to generate high resolution maps.
Satellite Images Can Pinpoint Poverty Where Surveys Can’t

Economic View
By SENDHIL MULLAINATHAN  APRIL 1, 2016
Ongoing work

- Describe, model, and predict **changes over time**

- Incorporate **new data sources** (phone data, crowdsourcing, etc.)

- Mapping and estimating **crop yields**
  - 1st prize at INFORMS yield prediction challenge

Credit: premise.com
Summary

• Introduction
• Machine Learning for Public Policy
• AI for Sustainable Energy
• Conclusion
Computational Sustainability

- Optimization
- Poverty traps
- Poverty mapping
- Natural resources management
- Optimization of energy systems
- Artificial Intelligence and Machine Learning
- Groundwater and weather systems modeling
- Large Datasets
- Energy Materials discovery

Artificial Intelligence and Machine Learning

Energy Materials discovery

Natural resources management

Optimization of energy systems

Groundwater and weather systems modeling

Large Datasets

Energy Materials discovery

Optimization

Poverty traps

Poor mapping

Natural resources management

Optimization of energy systems

Groundwater and weather systems modeling

Large Datasets

Energy Materials discovery
Goal
Accelerate the pace and reduce the cost of discovery, and deployment of advanced material systems

20 years → 5 years

Very exciting new research area for Computer Science and Big Data techniques
Vision: AI for materials research

Domain Knowledge → Experiment Design → High throughput experiments

Data analysis → Automatic Data Analysis
Slide courtesy of Apurva Mehta and Yijin Liu, SLAC

4 million XANES spectrums collected in a few minutes with 30 nm spatial resolution.
Pattern Decomposition with Complex Combinatorial Constraints: Application to Materials Discovery. [AAAI 2015]

Identify materials

Pixel(r) XANES

Pixel(b) XANES

NiO

Ni

NiO_{82\%} Ni_{18\%}

NiO_{23\%} Ni_{77\%}

Tomo scan, one chemical map per angle

1 Billion XANES spectra in 10 hours

3D structure
Vision: AI for materials research

Domain Knowledge → Experiment Design → High throughput experiments

Improved Data Collection

Data analysis

Stanford Linear Accelerator

Cornell High Energy Synchrotron Source

Energy Materials Center at Cornell

Caltech
Linac Coherent Light Source (LCLS) is the world's first X-ray laser. 10 billion times brighter than any other X-ray source before it.

Very complex machine, difficult to operate, requires manual tuning (hundreds of hours per year).

Operating cost close to $1,000 per minute – want to make parameter tuning as robust and as quick as possible.
Archiving system: records almost 200,000 independent variables once a second, and goes back several years

Bayesian optimization:

- Works by seeking promising points that aren’t already explored
- Sound way to deal with the classic exploration vs exploitation tradeoff

Sparse Gaussian Processes for Bayesian Optimization
[under review at UAI-16]
Vision: AI for materials research

Preliminary work on dielectric screening via quantum simulations

Domain Knowledge → Experiment Design → High throughput experiments → Data analysis

Cornell High Energy Synchrotron Source
Energy Materials Center at Cornell
Caltech
Stanford Linear Accelerator
Summary

• Introduction
• Machine Learning for Public Policy
• AI for Sustainable Energy
• Conclusion
Conclusions

• Growing concerns about the threats of Artificial Intelligence to the future of humanity

• Recent advances in AI also create enormous opportunities for having deeply beneficial influences on society (energy, sustainability, …)

• Exciting opportunities for Computer Science research