Machine Learning for Environmental Grand Challenges

Shakir Mohamed
Research Scientist, DeepMind

#AI4ER Cambridge
Statistical Operations

- Estimation and Learning
- Inference
- Comparison
- Data Enumeration
- Modelling
- Experimental Design
- Hypothesis Testing
Statistical Operations

Inference
What we can *know* about our data

Decision-making
What we can *do* with our data.
Part I: Pathways in Machine Learning

Shakir Mohamed
Research Scientist, DeepMind

#AI4ER Cambridge
On Models

Model: Description of the world, of data, of potential scenarios, of processes.

A probabilistic model writes out these models using the language of probability.

Probabilistic models let you learn probability distributions of data.

Most models in machine learning are probabilistic.
Statistical Inference

Learning Principles

Direct
- Laplace approximation
- Maximum a posteriori
- Cavity Methods
- Expectation Maximisation
- Noise Contrastive
- Maximum Likelihood
- Variational Inference
- Integr. Nested Laplace Approx
- Markov chain Monte Carlo
- Sequential Monte Carlo

Indirect
- Two Sample Comparison
- Approx Bayesian Computation
- Max Mean Discrepancy
- Method of Moments
- Transportation methods
Model Evidence

Model evidence (or marginal likelihood, partition function):
Integrating out any global and local variables enables model scoring, comparison, selection, moment estimation, normalisation, posterior computation and prediction.

Integral is intractable in general and requires approximation.

Learning principle: Model Evidence

\[ p(x) = \int p(x, z) dz \]

Basic idea: Transform the integral into an expectation over a simple, known distribution.
Variational Methods

Deterministic approximation procedures with bounds on probabilities of interest.

Fit the variational parameters.

$$KL[q(z|y) || p(z|y)]$$

True posterior

Approximation class $q_\phi(z)$
Learning by Comparison

We compare the estimated distribution \( q(x) \) to the true distribution \( p^*(x) \) using samples.

Basic idea: Transform into learning a model of the density ratio.

Learning principle: Two-sample tests

\[
\frac{p^*(x)}{q(x)} = 1 \quad p^*(x) = q(x)
\]

Interest is not in estimating the marginal probabilities, only in how they are related.
Estimation by Comparison

Two steps
1. Use a hypothesis test or comparison to obtain some model to tell how data from our model differs from observed data.

2. Adjust model to better match the data distribution using the comparison model from step 1.

Density Estimation by Comparison

Density Difference
\[ r_\phi = p^* - q_\theta \]

Density Ratio
\[ r_\phi = \frac{p^*}{q_\theta} \]

\[ H_0 : p^* = q_\theta \text{ vs. } p^* \neq q_\theta \]

Max Mean Discrepancy

Moment Matching

Bregman Divergence

Class Probability Estimation

\[ f(u) = u \log u - (u + 1) \log(u + 1) \]
Algorithms for Generative Models

Fully-observed autoregressive models

\[ p(x) \]

PixelCNN and Wavenet

Prescribed latent variable models and variational inference

\[ \tilde{p}(x) \leq p(x) \]

Variational Autoencoders

Implicit latent variable models and estimation-by-comparison

\[ r(x) = \frac{p^*(x)}{p(x)} \]

Generative Adversarial Networks
Common gradient problem

\[ \nabla \phi \mathbb{E}_{q_\phi(z)} [f_\theta(z)] = \nabla \int q_\phi(z) f_\theta(z) \, dz \]

1. **Pathwise estimator**: Differentiate the function \( f(z) \)
2. **Score-function estimator**: Differentiate the density \( q(z|x) \)

**Typical problem areas**
- Sensitivity analysis
- Generative models and inference
- Reinforcement learning and control
- Operations research and inventory control
- Monte Carlo simulation
- Finance and asset pricing
Progress in Generative Models

Figure 11: Generations of a DCGAN that was trained on the Imagenet-1k dataset.

Visual Quality of Independent Samples

Conceptual Compression

Figure 10. Generated samples from a network trained on 64×64 ImageNet with input scaling = 0.4. Qualitatively asking the model to be less precise seems to lead to visually more appealing samples.
Perception-Action Loops

**Biological perception-action loop**

- Environment
  - Observation/Sensation
  - Action
  - Primary Motor Cortex
  - Premotor Cortex
  - Prefrontal Cortex
  - Primary Sensory Cortex
  - Sensory Association Cortex
  - Posterior Assoc. Cortex

**Computational perception-action loop**

- External Environment
  - Observation/Sensation
  - Action
  - Internal Environment
    - Option KB
    - Critic
    - State Repr.
    - State Embedding
    - Planner

Shakir Mohamed
Action-conditional and latent-only transitions.

Grounded representations in actions and observations, using simulation to support grounding.
Intrinsic Motivation

Equip agents with mechanisms to produce and learn from internal rewards that can guide behaviour, when external rewards are absent.

\[ \mathcal{E}(s) = \max_{\omega} I^\omega(a, s'|s) \quad \max_{\omega} \mathbb{E}_{p(s'|a,s)\omega(a|s)} \left[ \log \frac{p(s', a|s)}{p(s'|s)\omega(a|s)} \right] \]
Mnih et al. (2015)
Generalising AlphaGo to any 2-player game

Fully general; No opening book; No endgame database; No heuristics; Starts from random

All learned **without any reference to past human games**
Applications in Healthcare

1. Better clinical outcomes
2. Enhance patient and clinician experience
3. Reduce costs

Data used by the model:
- 24h
- 48h
- 72h

Optional longer history:
- AKI Predicted
- 24h
- 48h
- 72h

Timeline:
- Outpatient events
- Admission
- Time unknown

Model
Predicting Organ Failure

Make predictions of AKI up to 48hr ahead. Provide a window for meaningful action.

For the most severe cases, can detect up to 90% of cases.

Tomasev et al. (2019)
Consider the uses of our models.
What are the dual uses of generative models. How do we think critically about these uses, educate, regulate, co-design these tools.
Dual-uses and Value Alignment
Neutrality and Universality

Neutrality Traps

• The Portability Trap: Failure to understand how repurposing algorithmic solutions designed for one social context may be inaccurate / do harm when applied to a different context.

• The Formalism Trap: Failure to account for the full meaning of social concepts such as fairness, which be resolved through mathematical formalisms.

• The Ripple Effect Trap: Failure to understand how the insertion of technology into an existing social system changes the behaviours and embedded values of the pre-existing system.

• The Solutionism Trap: Failure to recognise the possibility that the best solution to a problem may not involve technology.

Universality

‘A mono-cultural view of ethics conceives itself as the only valid one. In order to avoid this kind of ethical chauvinism and colonialism it is necessary that transcultural ethics arise from an intercultural dialogue instead of thinking of itself as universal without noticing its own cultural bias.’ Capurro, 2004
Part II: AI for Environmental Risk

Shakir Mohamed
Research Scientist, DeepMind

#AI4ER Cambridge
Extreme Weather Events

Segment Tropical Cyclones, Atmospheric Rivers from background

Given CAM5 outputs of a tropical cyclone and its initial position, track its trajectory.

Tools for data assimilation, analysis of NWP simulations, and new types of decision support.

Mudigonda et al. (2017)
Hybrid Physical Process Modelling

Predict future sea surface temperature (SST) from previous synthetic SST data from NEMO (Nucleus for European Modeling of the Ocean)

**Physical Model: Advection-Diffusion Equation**

\[
\frac{\partial I}{\partial t} + (w \cdot \nabla) I = D \nabla^2 I
\]

**Solution**

\[
I(x, t) = \int_{\mathbb{R}^2} k(x - w, y) I_0(y) \, dy
\]

\[
\hat{I}_{t+1}(x) = \sum_{y \in \Omega} k(x - \hat{w}(x), y) I_t(y)
\]

Key Idea: Predict \( w \)

\[
L_t = \sum_{x \in \Omega} \left\| \hat{I}_{t+1}(x) - I_{t+1}(x) \right\|^2 + \lambda_{\text{div}} \left\| \nabla \cdot \hat{w}_t(x) \right\|^2 + \lambda_{\text{magn}} \left\| \hat{w}_t(x) \right\|^2 + \lambda_{\text{grad}} \left\| \nabla \hat{w}_t(x) \right\|^2
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Average Score (MSE)</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical model [1]</td>
<td>1.99</td>
<td>4.8 s</td>
</tr>
<tr>
<td>ConvLSTM [9]</td>
<td>5.76</td>
<td>0.018 s</td>
</tr>
<tr>
<td>ACNN</td>
<td>15.84</td>
<td>0.54 s</td>
</tr>
<tr>
<td>GAN Video Generation ([7])</td>
<td>4.73</td>
<td>0.096 s</td>
</tr>
<tr>
<td>Proposed model with regularization</td>
<td>1.42</td>
<td>0.040 s</td>
</tr>
<tr>
<td>Proposed model without regularization</td>
<td>2.01</td>
<td>0.040 s</td>
</tr>
</tbody>
</table>
Solar Nowcasting

**Predict solar irradiance, accounting for clouds.**

- Numerical weather models become out of date with respect to the most recent observations.
- Solar irradiance is greatly affected by clouds; operational numerical weather models can’t resolve clouds.
- Radiative transfer codes in numerical weather models are some of the most computationally expensive bits of numerical weather models.

---

Shakir Mohamed

Kelly (2019), wikipedia.
Energy Consumption

Dramatically increase efficiency of existing systems
Application to Google Data Centres
Data centres across the world use around **3% of the world’s electricity**

Cooling energy is the largest non-server load (up to 40% of total energy usage)

**State**
- Incoming IT load
- Power meters
- Pressure sensors
- Temperature sensors
- Water flow meters
- Pump and fan speeds
- Fault alarms
- Weather conditions

**Actions**
- Number of cooling towers
- Number of chillers
- Number of pumps
- Temperature setpoints
- Pressure setpoints
- Flow setpoints
- Valve positions

Over 1,200 state variables and 20 actions
General Learning Framework for DC Operations

**State inputs**
- Current IT load
- Power meters
- Pressure sensors
- Temp sensors
- Weather
- Fan speeds
- ...

**Actions**
- # active coolers
- # chillers
- Pumps on/off
- Temp setpoints
- Valve setpoints
- Pressure setpoints
- ...

Gamble and Gao (2018)
Every five minutes: generate recommendations, send to a human operator for implementation
40% reduction in data center cooling energy
Spread the load across more equipment.
Local v. Global trade-offs.

Higher flow is not always better.
Reduced water flow to chillers in some weather conditions.

Shifting the loads.
Learned to shift cooling load to components that were more or less efficient at different times of year.
After three quarters of operation, scaling it up and getting it into production using a safety-first automation approach.

Recommendations are sent directly to the data centre, to be verified by the local controls system for safety before implementation.
Safety-first for direct AI control

- Continuous monitoring
- Automatic failover
- Smooth transfer
- Two-layer verification
- Constant communication
- Uncertainty estimation
- Rules and heuristics as backup
- Human in the loop

Gamble and Gao (2018)
Managing Energy Generation

Improving the economics of wind energy to accelerate adoption

The cost of turbines has plummeted, but wind is **unpredictable** and **intermittent**

The unpredictability of renewable energy makes it **less valuable** than fossil fuel energy

One strategy: train a system for **predicting** and **scheduling** wind energy
Applying ML algorithms to **700MW** of Google’s wind farm portfolio.
Inputs
- Global numerical weather forecasts
- Local weather observations

Outputs
- Future wind power output (36 hours in advance)

Wind Power: Predicted Output vs Ground Truth

Elkin and Witherspoon (2019)
20% increase in economic value, compared to baseline of no time-based commitments to grid
Tackling Climate Change with Machine Learning

David Rolnick1, Priya L. Doni2, Lynn H. Kaack3, Kelly Kochanski4, Alexandre Lacoste5, Kris Sankaran6, Andrew Slinin Ross6, Nikola Milojevic-Dupont6,10, Natasha Jaques11, Anna Waldman-Brown11, Alexandra Luccion11, Tegan Maharaj7, Evan D. Sherwin2, S. Karthik Mukkavilli6, Konrad P. Kording1, Carla Gomes12, Andrew Y. Ng13, Demis Hassabis14, John C. Platt15, Felix Creutzig9,16, Jennifer Chayes15, Yoshua Bengio5,7

1University of Pennsylvania, 2Carnegie Mellon University, 3ETH Zürich, 4University of Colorado Boulder, 5Element AI, 6Mila, 7Université de Montréal, 8Harvard University, 9McMaster Research Institute on Global Commons and Climate Change, 10Technische Universität Berlin, 11Massachusetts Institute of Technology, 12Cornell University, 13Stanford University, 14DeepMind, 15Google AI, 16Microsoft Research

Abstract

Climate change is one of the greatest challenges facing humanity, and we, as machine learning experts, may wonder how we can help. Here we describe how machine learning can be a powerful tool in reducing greenhouse gas emissions and helping society adapt to a changing climate. From smart grids to disaster management, we identify high impact problems where existing gaps can be filled by machine learning, in collaboration with other fields. Our recommendations encompass exciting research ques-

CHAPTER 4

Climate Informatics

Claire Monteleoni, Gavin A. Schmidt, Francis Alexander, Alexandru Niculescu-Mizil, Karsten Steinhaeuser, Michael Tippett, Arindam Banerjee, M. Benno Blumenthal, Auroop R. Ganguly, Jason E. Smerdon, and Marco Tedesco
References

- de Bézenac, E., Pajot, A., & Gallinari, P. Towards a Hybrid Approach to Physical Process Modeling.
- C. Gamble and J. Gao. Safety-first AI for autonomous data centre cooling and industrial control
Machine Learning for Environmental Grand Challenges

With thanks to colleagues and the work of many others referenced here from our ML community.

Shakir Mohamed
Research Scientist, DeepMind

#AI4ER Cambridge