

Automating Science with Artificial Intelligence

Gilles Louppe, Faculty of Applied Sciences g.louppe@uliege.be @glouppe Looking for **new** PhD students! (Computer Science, Engineering or Physical Sciences)

Joint work with Joeri Hermans (ULiège), Kyle Cranmer (NYU), Lukas Heinrich (NYU), David Rousseau (LAL)



2. Compute the consequences of the guess

3. Compare the results to Nature

Long-term research goal: automate the Scientific method through modern Artificial Intelligence and Deep Learning methods!

 $\begin{aligned} \mathcal{L}_{GWS} &= \sum_{f} (\bar{\Psi}_{f} (i\gamma^{\mu} \partial \mu - m_{f}) \Psi_{f} - eQ_{f} \bar{\Psi}_{f} \gamma^{\mu} \Psi_{f} A_{\mu}) + \\ &+ \frac{g}{\sqrt{2}} \sum_{i} (\bar{a}_{L}^{i} \gamma^{\mu} b_{L}^{i} W_{\mu}^{+} + \bar{b}_{L}^{i} \gamma^{\mu} a_{L}^{i} W_{\mu}^{-}) + \frac{g}{2c_{w}} \sum_{f} \bar{\Psi}_{f} \gamma^{\mu} (I_{f}^{3} - 2s_{w}^{2} Q_{f} - I_{f}^{3} \gamma_{5}) \Psi_{f} Z_{\mu} + \\ &- \frac{1}{4} |\partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu} - ie(W_{\mu}^{-} W_{\nu}^{+} - W_{\mu}^{+} W_{\nu}^{-})|^{2} - \frac{1}{2} |\partial_{\mu} W_{\nu}^{+} - \partial_{\nu} W_{\mu}^{+} + \\ &- ie(W_{\mu}^{+} A_{\nu} - W_{\nu}^{+} A_{\mu}) + ig' c_{w} (W_{\mu}^{+} Z_{\nu} - W_{\nu}^{+} Z_{\mu}|^{2} + \\ &- \frac{1}{4} |\partial_{\mu} Z_{\nu} - \partial_{\nu} Z_{\mu} + ig' c_{w} (W_{\mu}^{-} W_{\nu}^{+} - W_{\mu}^{+} W_{\nu}^{-})|^{2} + \\ &- \frac{1}{2} M_{\eta}^{2} \eta^{2} - \frac{g M_{\eta}^{2}}{2M_{\eta}} \eta^{3} - \frac{g'^{2} M_{\eta}^{2}}{2(2M_{\eta}^{-})} \eta^{4} + |M_{W} W_{\mu}^{+} + \frac{g}{2} \eta W_{\mu}^{+}|^{2} + \end{aligned}$

II. Optimal experimental design

How do we design the best scientific experiment?

If an experiment is configured by ϕ , then we want to find the experiment that maximizes the reduction in entropy $\Delta(\phi)$, therefore obtaining the least amount of uncertainty H[θ].

Solution: optimal experimental design!





Standard Model of Particle Physics

I. Likelihood-free inference

To validate a guess and compare it to Nature, we need to compute its consequences. The expected consequences are usually modelled through a computer simulator.



ters

In n dura invo adm

Whi

Issue: computation of $p(x | \theta)$ is in most cases intractable.

Solutions?

Approximating Likelihood Ratios with Calibrated Discriminative Classifiers

Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe¹ and Kyle Cranmer¹ ¹New York University

Challenges:

- infrastructure: full reproducible workflow needs to be run multiple times in order to make the estimations for different values of ϕ .
- optimization: estimating the reduction of uncertainty is slow. We need an efficient optimization algorithm (e.g., Bayesian optimization).

III. Exploring the theory space

Can we automatically explore the space of theories and find the envelope that

 $MG + G(GM^T + G) + G$

agrees with the data?

Solutions?

- Symbolic exploration —
- Finding exclusion contours, while minimizing the number of calls (expensive): $\{\psi | \rho(p_r(\mathbf{x} | \phi), p(\mathbf{x} | \psi, \phi, \theta^*)) < \epsilon\}$



Kyle Cranmer¹, Juan Pavez², and Gilles Louppe¹ ¹New York University ²Federico Santa María University

March 21, 2016

Abstract

In many fields of science, generalized likelihood ratio tests are established tools for statistical inference. At the same time, it has become increasingly common that a simulator (or generative model) is used to describe complex processes that tie parameters θ of an underlying theory and measurement apparatus to high-dimensional observations $\mathbf{x} \in \mathbb{R}^p$. However, simulator often do not provide a way to evaluate the likelihood function for a given observation \mathbf{x} , which motivates a new class of likelihood-free inference algorithms. In this paper, we show that likelihood ratios are invariant under a specific class of dimensionality reduction maps $\mathbb{R}^p \mapsto \mathbb{R}$. As a direct consequence, we show that discriminative classifiers can be used to approximate the generalized likelihood ratio statistic when only a generative model for the data is available. This leads to a new machine learning-based approach to likelihood-free inference that is complementary to Approximate Bayesian Computation, and which does not require a prior on the model parameters. Experimental results on artificial problems with known exact likelihoods illustrate the potential of the proposed method.

Keywords: likelihood ratio, likelihood-free inference, classification, particle physics, surrogate model Complex computer simulators are increasingly used across fields of science as generative models tying parameters of an underlying theory to experimental observations. Inference in this setup is often difficult, as simulators rarely admit a tractable density or likelihood function. We introduce Adversarial Variational Optimization (AVO), a likelihood-free inference algorithm for fitting a nondifferentiable generative model incorporating ideas from empirical Bayes and variational inference. We adapt the training procedure of generative adversarial networks by replacing the differentiable generative network with a domain-specific simulator. We solve the resulting non-differentiable minimax problem by minimizing variational upper bounds of the two adversarial objectives. Effectively, the procedure results in learning a proposal distribution over simulator parameters, such that the corresponding marginal distribution of the generated data matches the observations. We present results of the method with simulators producing both discrete and continuous data.

I. INTRODUCTION	II. PROBLEM STATEMENT
many fields of science such as particle physics, epi- ology, and population genetics, computer simulators sed to describe complex data generation processes. e simulators relate observations \mathbf{x} to the parame- $\boldsymbol{\theta}$ of an underlying theory or mechanistic model. ost cases, these simulators are specified as proce- implementations of forward, stochastic processes ving latent variables \mathbf{z} . Rarely do these simulators \mathbf{t} a tractable density (or likelihood) $p(\mathbf{x} \boldsymbol{\theta})$. The lence and significance of this problem has motivated tive research effort in so-called <i>likelihood-free infer</i> - algorithms such as Approximate Bayesian Compu- n (ABC) and density estimation-by-comparison al- ums [1–6]. parallel, with the introduction of variational auto- lers [7] and generative adversarial networks [8],	We consider a family of parametrized densities $p(\mathbf{x} \boldsymbol{\theta})$ defined implicitly through the simulation of a stochas- tic generative process, where $\mathbf{x} \in \mathbb{R}^d$ is the data and $\boldsymbol{\theta}$ are the parameters of interest. The simulation may in- volve some complicated latent process where $\mathbf{z} \in \boldsymbol{Z}$ is a latent variable providing an external source of random- ness. Unlike implicit generative models defined by neural networks, we do not assume \mathbf{z} to be a fixed-size vector with a simple density. Instead, the dimension of \mathbf{z} and the nature of its components (uniform, normal, discrete, continuous, etc.) are inherited from the control flow of the simulation code and may depend on $\boldsymbol{\theta}$ in some in- tricate way. Moreover, the dimension of \mathbf{z} may be much larger than the dimension of \mathbf{x} . We assume that the stochastic generative process that defines $p(\mathbf{x} \boldsymbol{\theta})$ is specified through a non-differentiable deterministic function $g(\cdot; \boldsymbol{\theta}) : \boldsymbol{Z} \to \mathbb{R}^d$. Operationally,
has been a vibrant research program around im- generative models based on neural networks [9]. a some of these models also do not admit a tractable	$\mathbf{x} \sim p(\mathbf{x} \boldsymbol{\theta}) \equiv \mathbf{z} \sim p(\mathbf{z} \boldsymbol{\theta}), \mathbf{x} = g(\mathbf{z}; \boldsymbol{\theta})$ (1) such that the density $p(\mathbf{x} \boldsymbol{\theta})$ can be written as
ty, they are all differentiable by construction. In ad- a, generative models based on neural networks are a parametrized and the model parameters have no	$p(\mathbf{x} \boldsymbol{\theta}) = \int_{\{\mathbf{z}:g(\mathbf{z};\boldsymbol{\theta})=\mathbf{x}\}} p(\mathbf{z} \boldsymbol{\theta})\mu(d\mathbf{z}), \tag{2}$
us interpretation. In contrast, scientific simulators e thought of as highly regularized generative mod- they typically have relatively few parameters and	where μ is a probability measure. Given some observed data $\{\mathbf{x}_i i = 1,, N\}$ drawn from the (unknown) true distribution $p_r(\mathbf{x})$, our goal is to

they are endowed with some level of interpretation. In this setting, inference on the model parameters θ is often of more interest than the latent variables z. In this note, we develop a likelihood-free inference algorithm for non-differentiable, implicit generative models. We adapt the adversarial training procedure of generative adversarial networks [8] by replacing the implicit generative network with a domain-based scientific simulator, and solve the resulting non-differentiable minimax define the parameters θ that minimize the divergence between $p_r(\mathbf{x})$ and the implicit model $p(\mathbf{x}|\theta)$. That is, $\theta^* = \arg \min_{\theta} \rho(p_r(\mathbf{x}), p(\mathbf{x}|\theta)),$ (if where ρ is some distance or divergence. III. BACKGROUND

Adversarial Variational Optimization (Louppe, Cranmer, 2017):

- Same as generative adversarial networks (see below).
- Replace generator network with domain simulator.
- Cope with non-differentiability using variational optimization.
- Require state-of-the-art classifiers for optimal performance.

Deep learning for physical sciences strongly needed here!

Generative adversarial networks:





0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 0.00 0.25 0.50 0.75 1.00 1.25 1.5

IV. Use cases (ongoing work)

Particle physics (collaborations with CERN and Oxford)

- Exploration of Beyond the Standard Model theories
- Automated tuning of particle detectors





Cosmology (collaboration with astrophysicists)



Kyle Cranmer

New York University

kyle.cranmer@nyu.edu

Geomorphology (collaboration with Earth scientists)



Which of those are real faces?

V. Deep learning for physical sciences

This interdisciplinary research usually requires domain-specific deep learning models, which results in contributions of their own.

Implicit models: Likehood-free inference relates to a large body of statistical methods developed within the machine learning community:

- Approximate Bayesian Computations
- Density estimation-by-comparisonVaritional inference



Learning to Pivot with Adversarial Networks

Michael Kagan

makagan@slac.stanford.edu

Several techniques for domain adaptation have been proposed to account for

differences in the distribution of the data used for training and testing. The majority

of this work focuses on a binary domain label. Similar problems occur in a scientific

context where there may be a continuous family of plausible data generation

processes associated to the presence of systematic uncertainties. Robust inference is possible if it is based on a pivot - a quantity whose distribution does not depend on the unknown values of the nuisance parameters that parametrize this family

of data generation processes. In this work, we introduce and derive theoretical

results for a training procedure based on adversarial networks for enforcing the

pivotal property (or, equivalently, fairness with respect to continuous attributes) on

a predictive model. The method includes a hyperparameter to control the trade-

off between accuracy and robustness. We demonstrate the effectiveness of this

approach with a toy example and examples from particle physics.

New York University SLAC National Accelerator Laboratory

Gilles Louppe

g.louppe@nyu.edu

Neural Message Passing for Jet Physics

Isaac Henrion, Johann Brehmer, Joan Bruna, Kyunghun Cho, Kyle Cranmer Center for Data Science New York University New York, NY 10012 {henrion, johann.brehmer, bruna, kyunghyun, kyle.cranmer}@nyu.edu Gilles Louppe Department of Computer Science University Computer Science

University of Liège Belgium g.louppe@ulg.ac.be University of Liège Belgium g.louppe@ulg.ac.be Belgium gaspar.rochette@ens.fr

Abstract

Supervised learning has incredible potential for particle physics, and one application that has received a great deal of attention involves collimated sprays of particles called jets. Recent progress for jet physics has leveraged machine learning techniques based on computer vision and natural language processing. In this work, we consider message passing on a graph where the nodes are the particles in a jet. We design variants of a message-passing neural network (MPNN); (1) with a learnable adjacency matrix, (2) with a learnable symmetric adjacency matrix, and (3) with a set2set aggregated hidden state and MPNN with an identity adjacency matrix. We compare these against the previously proposed recursive neural network with a fixed tree structure and show that the MPNN with a learnable adjacency matrix and two message-passing iterations outperforms all the others.

QCD-Aware Recursive Neural Networks for Jet Physics

Gilles Louppe,¹ Kyunghyun Cho,¹ Cyril Becot,¹ and Kyle Cranmer¹ ¹New York University

Recent progress in applying machine learning for jet physics has been built upon an analogy between calorimeters and images. In this work, we present a novel class of recursive neural networks built instead upon an analogy between QCD and natural languages. In the analogy, four-momenta are like words and the clustering history of sequential recombination jet algorithms is like the parsing of a sentence. Our approach works directly with the four-momenta of a variable-length set of particles, and the jet-based tree structure varies on an event-by-event basis. Our experiments highlight the flexibility of our method for building task-specific jet embeddings and show that recursive architectures are significantly more accurate and data efficient than previous image-based networks. We extend the analogy from individual jets (sentences) to full events (paragraphs), and show for the first time an event-level classifier operating on all the stable particles produced in an LHC event.