Through the Eyes of Birds and Frogs

WRITING AND SURVEYS IN MACHINE LEARNING RESEARCH

SHAKIR MOHAMED

@SHAKIR_ZA
Birds fly high in the air and *survey broad vistas* of mathematics out to the far horizon. They delight in concepts that unify our thinking.

Frogs live in the mud below and see only the flowers that grow nearby. They *delight in the details* of particular objects.

—Freeman Dyson (2009), *Birds and Frogs.*
To aim in our writing to familiarise the strange and to mystify the familiar.

-Toni Morrison
Mystifying and Familiarising

All our work contains an element of survey and review: all our papers, blogs, essays, reviews, theses, perspectives, books.

Thinking of surveys is instructive, and applies to the smallest section, all the way to an entire paper.
A REVIEW OF RADAR-BASED NOWCASTING OF PRECIPITATION AND APPLICABLE MACHINE LEARNING TECHNIQUES

Rachel Prudden1,†, Samantha Adams1, Dmitry Kangin1,†, Niall Robinson1,†, Sumran Ravari†,†, Shakir Mohamed1,∗, Alberto Arribas1,†

ABSTRACT

A ‘nowcast’ is a type of weather forecast which makes predictions in the very short term, typically less than two hours – a period in which traditional numerical weather prediction can be limited. This type of weather prediction has important applications for commercial aviation, public and outdoor events; and the construction industry, power utilities, and ground transportation services that conduct much of their work outdoors. Importantly, one of the key needs for nowcasting systems is in the provision of accurate warnings of adverse weather events, such as heavy rain and flooding, for the protection of life and property in such situations. Typical nowcasting approaches are based on simple extrapolation models applied to observations, primarily rainfall radar. In this paper we review existing techniques to radar-based nowcasting from environmental sciences, as well as the statistical approaches that are applicable from the field of machine learning. Nowcasting continues to be an important component of operational systems and we believe new advances are possible with new partnerships between the environmental science and machine learning communities.

Keywords: Monte Carlo gradient estimation, sensitivity analysis, score function estimator, pathwise estimator, measure-valued gradient estimators, variance reduction.©}

Case Studies from my Own Writing

Monte Carlo Gradient Estimation in Machine Learning

Shakir Mohamed†,∗, Milaëla Ronco†,‡, Michael Figurnov†,‡, Andry Mail†,‡

∗Equal contributions: 1 DeepMind, London 2 University College London

Editor: Jon McAuliffe

Abstract

This paper is a broad and accessible survey of the methods we have at our disposal for Monte Carlo gradient estimation in machine learning and across the scientific sciences: the problem of computing the gradient of an expectation of a function with respect to parameters defining the distribution that is integrated; the problem of sensitivity analysis. In machine learning research, this gradient problem lies at the core of many learning problems, in supervised, unsupervised and reinforcement learning. We will generally seek to rewrite such gradients in a form that allows for Monte Carlo estimation, allowing them to be easily and efficiently used and analysed. We explore three strategies—the pathwise, score function, and measure-valued gradient estimators—exploring their historical development, derivation, and underlying assumptions. We describe their use in other fields, show how they are related and can be combined, and expand on their possible generalisations. Wherever Monte Carlo gradient estimators have been derived and deployed in the past, important advances have followed. A deeper and more widely-held understanding of this problem will lead to further advances, and it is these advances that we wish to support.

Keywords: gradient estimation, Monte Carlo, sensitivity analysis, score function estimator, pathwise estimator, measure-valued estimator, variance reduction.
Monte Carlo Gradient Estimation

Fundamental problem in computational sciences: compute the gradient of an expectation.

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

Widely known as sensitivity analysis, but appears in many fields. In machine learning alone this problem is central to learning in generative models, reinforcement learning, Bayesian optimisation, and active learning.
Monte Carlo Gradient Estimation

Structure of this paper:
1. Stochastic optimisation and sensitivity analysis.
2. Monte Carlo methods
3. Intuitive understanding of methods.
4. Three estimators: Score function, pathwise, and measure-valued gradient estimators.
   - Basic tools, deriving the estimator, estimator variance, properties, computation
5. Experiments
6. Other estimators and connections
7. Conclusions
Geometric Forms of Narrative

In the study of writing we will often find recurring patterns that are described using geometric imagery.

Inverted Pyramid

Spiral
Zinsser’s Guidance

Move beyond fact into significance and speculation: how the exposition alters what was known, what new avenues of research it might open, where the research might be applied.

Broaden the first point.

One fact a reader must know
Spiral Narrative

Create flow and structure by using repeated patterns.

In section 3-5 of the MC gradients paper, we use the same structure for every section: can read every section in parallel and you (hopefully) learn the same types of tools at the same pace, and reveals the logic of the analysis. Same headings, proof questions, types of conclusions.
Radar-based Nowcasting of Precipitation

Challenges in writing between different fields that have long-established traditions, philosophy and practice.

Structure
- Motivating the applied problem.
- Physics of precipitation.
- Physics-derived models.
- Data-driven models.
- Future for mixed methods
- Glossary.

A REVIEW OF RADAR-BASED NOWCASTING OF PRECIPITATION AND APPLICABLE MACHINE LEARNING TECHNIQUES

Abstract
A ‘nowcast’ is a type of weather forecast which makes predictions in the very short term, typically less than two hours - a period in which traditional numerical weather prediction can be limited. This type of weather prediction has important applications for commercial aviation, public and outdoor events, and the construction industry, power utilities, and ground transportation services that conduct much of their work outdoors. Importantly, one of the key needs for nowcasting systems is in the provision of accurate warnings of adverse weather events, such as heavy rain and flooding, for the protection of life and property in such situations. Typical nowcasting approaches are based on simple extrapolation models applied to observations, primarily rainfall radar. In this paper we review existing techniques to radar-based nowcasting from environmental sciences, as well as the statistical approaches that are applicable from the field of machine learning. Nowcasting continues to be an important component of operational systems and we believe new advances are possible with new partnerships between the environmental science and machine learning communities.
Other Factors

Collaboration

Scope and Length

Take Time

Readers
Blogs

Our field has a strong culture of blogging.

Blogs free us to explore new ways of familiarising and mystifying.

Consider blogs in languages other than English as well.
Some Other Nice Reviews

How Computer Systems Embody Values

Helen Nissenbaum, Princeton University

Training as a philosopher has never been increasingly drawn toward the science and engineering of information technology in my work on its ethical, social, and political dimensions. I trace this interest back to a research project on computer systems’ bias with Ira N. Friedenberg (“Bias in Computer Systems,” ACM Trans. Information Systems, July 1994, pp. 340–46).

Focusing on social changes...

In one category 1 place work in which values themselves are not the controversy's central thesis. Thus, when researchers worry about computer systems replacing humans who act in positions of responsibility—planning, drug making, investment decisions, controlling aircraft, and so on—they do not

Variational Inference: A Review for Statisticians

David M. Blei¹, Alp Kucukelbir², and Jon D. McAuliffe²
¹Department of Computer Science and Statistics, Columbia University, New York, NY. ²Department of Computer Science, Columbia University, New York, NY. Department of Statistics, University of California, Berkeley, CA

ABSTRACT

One of the core problems of modern statistics is to approximate difficult-to-compute probability densities. This problem is especially important in Bayesian statistics, which frames all inference about unknown quantities as a calculation involving the posterior density. In this article, we review variational inference (VI), a method from machine learning that approximates probability densities through optimization. VI has been used in many applications and tends to be faster than classical methods, such as Markov chain Monte Carlo sampling. The idea behind VI is to first post a family of densities and then to find a member of that family which is close to the target density. Closeness is measured by Kullback–Leibler divergence. We review the ideas behind mean-field variational inference, discuss the special case of VI applied to exponential family models, present a full example with a Bayesian mixture of Gaussians, and derive a variant that uses stochastic optimization to scale up to massive data. We discuss modern research in VI and highlight important open problems. VI is powerful, but it is not yet well understood. Our hope in writing this article is to catalyze statistical research on this class of algorithms. Supplementary materials for this article are available online.
Shitty First Drafts
Anne Lamott from *Bird by Bird*

Born in San Francisco in 1954, Anne Lamott is a graduate of Goucher College in Baltimore and is the author of six novels, including Rosie (1983), Crooked Little Heart (1997), All New People (2000), and Blue Shoes (2002). She has also been the food reviewer for California magazine, a book reviewer for Mademoiselle, and a regular contributor to Salon's “Mothers Who Think.” Her nonfiction books include Operating Instructions: A Journal of My Son’s First Year (1993), in which she describes her adventures as a single parent, and Tender Mercies: Some Thoughts on Faith (1999), in which she charts her journey toward faith in God.

In the following selection, taken from Lamott’s popular book about writing, *Bird by Bird* (1994), she argues for the need to let go and write those “shitty first drafts” that lead to clarity and sometimes brilliance in our second and third drafts.

Now, practically even better news than that of short assignments is the idea of shitty first drafts. All good writers write them. This is how they end up with good second drafts and terrific third drafts. People tend to look at successful writers who are getting their books published and maybe even doing well financially and think that they sit down at their desks every morning feeling like a million dollars, feeling great about who they are and how much talent they have and what a great story they have to tell; that they take in a few deep breaths, push back their sleeves, roll their necks a few times to get all the cricks out, and dive in, typing fully formed passages as fast as a court reporter. But this is just the fantasy of the uninitiated. I know some very great writers, writers you love who write beautifully and have made a great deal of money, and not one of them sits down routinely feeling wildly enthusiastic and confident. Not one of them writes elegant first drafts. All right, one of them does, but we do not like her very much. We do not think that she has a rich inner life or that God likes her or can even stand her. (Although when I mentioned this to my priest friend Tom, he said you can safely assume you’ve created God in your own image when it turns out that God hates all the same people you do.)

The only way I can get anything written at all is to write really, really shitty first drafts.

—Anne Lamott
To aim in our writing to “...come across as people: men and women finding a common thread of humanity between themselves and their speciality and their readers.”

-William Zinsser
Through the Eyes of Birds and Frogs
Writing and Surveys in Machine Learning Research

SHAKIR MOHAMED
@SHAKIR_ZA
Some Resources