Micropediction

-1
0
+1

Microprediction

Building an Open AI Network

Peter Cotton

The MIT Press
Cambridge, Massachusetts
London, England

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Sabon by Westchester Publishing Services. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data is available.

ISBN:

10 9 8 7 6 5 4 3 2 1

To my mother, who taught me to question pyramids.

To know of and put to use a machine not fully employed, or somebody’s skill which could be better utilized, or to be aware of a surplus stock which can be drawn upon during an interruption of supplies, is socially quite as useful as the knowledge of better alternative techniques.

—Hayek (1945), *The Use of Knowledge in Society*

Contents

-1
0
+1

The pitch

Quants are meek. This is good because we will inherit the Earth. It's not so good when it comes to influencing the perception of quantitative work, or its future—the topic of this book.

These days everybody wants to be a data scientist. I get it, but I'm not heavily vested in vague terminology like data science or artificial intelligence. In these pages I set out to effect a qualitative change in your perception of whatever-you-want-to-call-it, and where it is going. I'm feel no obligation to defend any of the edifice that has arisen in recent years.

Ironically, I am asymptotically the world's most productive data scientist—as you will eventually discover—and unlike most, I'm prepared to tell you that most "new" activity can be broken down into standard, commodity, repeated quantitative tasks and delivered at much, much lower cost.

(If I may be colloquial, data science is a rip-off).

These repeated tasks I speak of go by many names in industry, which is part of the problem. In this book I mostly use only one word: microprediction. That is the act of making thousands or millions of predictions of the same type over and over again.

This book explores the nature of microprediction from the perspective of economics, statistics, decision making under uncertainty, and privacy preserving computation.

Informed by progress in those fields, I ask basic questions. What is the best way to produce and distribute high quality microprediction at arbitrarily low cost, and thus help businesses of all sizes? Is microprediction an individual or a collective activity? What things, currently described as AI, can't be decomposed into microprediction?

It is very hard to reconcile the answers with artisan data science. Instead, the conclusion reached is that the world is missing a public utility, and companies are missing an important abstraction in their strategies which might enable them to use it.

That utility is a vast collection of live streams of data, inhabited by autonomous self-navigating models that crawl from one to the next and make predictions. I invite you to help me create it because the world is not short of algorithms. I conclude it is short of micro-managers of algorithms, as I refer to them, who are autonomous algorithm chauffeurs.

If an algorithm can drive a car from New York to San Francisco, it ought to be able to find its way from a code repository to a business problem without human intervention. The main difficulty is a the failure to reduce business problems to a combination of canonical microprediction tasks.

This book is aimed at *all* potential contributors to a prediction web, as it might be described. You can supply code, or launch algorithms, or create new feeds, or supply mathematical insight, or help in any number of technical ways.

You can also help by socializing the idea, or adding to the demand for *explicit* versus implicit microprediction—by nudging your data scientists to send their model residuals to their nearest microprediction "oracle". Five lines of Python won't kill anyone.

However you choose to come at this topic, I hope that some of you will come to see the prediction network as a long felt unmet need. A microprediction micro-economy is not an easy thing to bring about. It might take time. It may never reach critical scale without happenstance. Perhaps your picking up this book, or searching "microprediction" to find working prototypes, is precisely what is required.

1 Overview

Imagine awaking one day to a world in which every aspect of every business, large or small, is quantitatively optimized to the n'th degree. Imagine that this this optimization derives its strength, ultimately, from advanced mathematical technique applied to decision making under uncertainty, and that this is informed by trillions of data points drawing in every conceivable quantity that might reasonably be deemed relevant to the task at hand.

In this scenario every activity is optimized, not just the production line of companies that can afford to hire data scientists and lay down expensive analytics infrastructure. Every small business benefits, including the guy selling beer and peanuts at the baseball game. Every individual benefits and cost is negligible. A democratic AI miracle has occurred.

This book asks only one hypothetical question. How did this come about? Among all possible explanations for this seemingly unlikely outcome, we set out to determine the one with the highest probability.

1.1 Half of Statistics

We will have some minor terminological challenges that are entirely other peoples' fault. So I must be crystal clear about one thing—only half an AI miracle is contemplated.

I consider the use of applied statistics to the optimization of relatively fast moving operational problems only. We shall assume the instrumentation of those processes throws off sufficient data for so-called data-hungry methods to work well, or at least enough data to mechanically assess competing approaches. We shall assume that quantitative challenges *of the same kind* are thrown up over and over again—thousands or millions of times.

Methods termed *machine learning* are therefore bound to play a key role. There isn't a mathematically useful delineation of that field from statistics, or parts of applied mathematics for that matter, but one finds oneself trading accuracy for brevity, and slipping into the use of *machine learning* to describe a collection of methods that work surprisingly well when data is plentiful.

That's not entirely fair to any tribe, and the related schism in the field of statistics is discussed with marginally more nuance in Chapter ??. But my point is not to assert superiority of machine learning over inferential statistics, or anything else. Rather, in the fine mathematical tradition, I merely wish to simplify my task to one that can be solved.

Rather than tripping on terminological sensitivities I prefer the vulgarity of refering to my domain of problems as "half of statistics", though I'll let the reader decide what fraction of "applied something" really constitutes my scope. You may decide it is very close to zero, or very close to one.

Arguing for the latter, so many things will be instrumented in the future that it's tempting to say that data-rich problems are the bigger half—though of course I mean that in a short-term, commercial sense (since human survival or extinction should probably appear on the scales as well).

Arguing for the former, the inherent statistical difficulty of a task approaches zero in the limit of infinite data, assuming that past is prologue. All you need is a nearest neighbour search.

In a somewhat related vein, press coverage of artificial intelligence breakthroughs has reinforced the idea, pun intended, that bespoke statistical modeling can be discarded since algorithms can learn in model-free ways.

In *Reward is Enough*, a recent essay by pioneers of reinforcement learning Silver, Singh, Precup and Sutton, it is suggested that the objective of maximising reward is "enough to drive behaviour that exhibits most if not all attributes of intelligence that are studied in natural and artificial intelligence, including knowledge, learning, perception, social intelligence, language and generalisation."[1]

My goal is more modest. I use the term microprediction in an attempt to avoid any impression otherwise. I emphasize the problem domain (as distinct from which methods might work) and hopefully the term microprediction can de-anchor some readers from unhelpful connotations of one-off events. That said, we are most certainly talking about prediction.

Semantics aside, in Chapter ?? we'll allow our mind's eye to pan across a world of business problems and how they can be transformed with a sprinkling of *frequently repeated* prediction pixie dust. Suffice to say that

-1
0
+1

the larger that opportunity is, the more important our contemplations will be. And the domain is growing rapidly.

Carrying forward the ambiguity in the word "half", I'll boldly assert that half of statistics—so defined—is half of AI. Part of the case rests on Chapter ?? where we consider frequently repeated conditional prediction of so-called value functions.

What is important, I feel, is to focus solely on frequently repeated prediction and data-rich problems, and not covet the other half of statistics.

I don't want the world
I just want your half
—Ana Ng, *They Might be Giants*

It is my assertion that companies have failed, in large part, to properly delineate the two, and reorganize accordingly.

1.2 Cost

Now that we know we are concerned with only half an AI democratic miracle, I'll move from the delineation of the problem domain to the assessment of quality.

A simple way to set the standard is to propose that *almost everything* will be predicted almost as well as *anything* is today. If that is our ambition, then cost is the only reason the miracle I describe can seemingly be reduced to absurdity. Cost *is* the problem.

Let's examine a small part of today's analytic world where no expense is spared. We view a hugely profitable market making firm. It is well positioned to hire the best and brightest minds it can throw money at. It burns through a ten figure technology budget, year after year, and leans on a virtually unlimited ability to store and process data.

This trading firm possesses a high level of organizational mathematical maturity, almost certainly represented in the technical acumen of senior management and flowing all the way down. It trains formidable intellectual firepower at a very precise and narrow task: determining the probabilistic representation of the near future of a finite set of security prices.

The firm does this because accuracy means profit. In Chapters ?? and ?? I'll give formal examples of that direct connection but for now, I ask you not to fixate on the details of any given task where accuracy helps but rather, how implausible it is that most ordinary businesses, including small ones, could achieve as high a level of predictive proficiency as this well funded organization.

I ask you to consider the enormity of the expense, and the incredible undertakings of this trading company over many years. No doubt this has consumed the best intellectual years of a large interdisciplinary team. All this energy, ultimately, is dedicated to the creation and maintenance of a continuously changing probabilistic description of a set of live numbers.

Those numbers are of economic significance it must be said—but nonetheless they are a tiny, mostly unrepresentative sample of the set of all quantities that instrument our personal and commercial lives. (It is possible, but unlikely, that a single number under study by this firm will materially help a surgeon, or a pilot, or an operator at a nuclear plant make a crucial, real-time decision.)

1.7 Orchestrating Prediction

Thus begins the clash of civilizations.

To create the best repeated short term prediction at a low cost, should humans organize in a large, well funded automated machine learning company? Or can the production of automated model selection be drawn together by other mechanisms? Rather obviously, loose collections of volunteers working on open source software also advance the needle.

But in this work, I focus on the price mechanism as a much more fine-grained orchestration principle, and I draw attention to the severe limitations of central planning. My response to Domingo’s invitation to create the master algorithm is to hope that we are already living in one—the economy. It’s merely that this master algorithm is far too cumbersome to achieve our microprediction objective, unless we change it.

1.12 Summary and Outline

In an imagined future where ubiquitous real-time operational intelligence has arrived at zero cost, I have prompted the reader to speculate as to the most likely origin.

Maybe this all starts when a hyper-intelligent agent with generalized intelligence escapes from the DeepMind lab and starts scanning the paper for data science job openings. That might even leave industry largely unchanged, from the organizational perspective.

However I've suggested a rather different possibility: a world-wide prediction web analogous in some ways to the internet itself. I offer the following line of argument.

1. Most real-time operational optimization can be formulated in terms of frequently repeated predictions of instrumented quantities, intermediate rewards, differences of value functions (Chapter ??) averages of predictions yet received, or something else.
2. We interpret the existence of the machine learning revolution as a statement that most models can be assessed in mechanical fashion.
3. Therefore only trade friction prevents reward from being enough (for repeated prediction). Only trade friction prevents the emergence of radically low cost self-organizing supply chains for microprediction.
4. When algorithms can traverse to repeated statistical games, and businesses abstract away microprediction from the rest of the applications logic, direct search cost plummets. The lemons problem fades also (due to the Law of Large Numbers). And the walls that separate us (privacy, intellectual property) can be addressed by skullduggery of various kinds (Chapter ??).
5. Microprediction quality can also continuously improve as costs fall due to network feedback (more data, more models, shared feature spaces).
6. At these extremely low levels of economic friction and microprediction cost, further feedback occurs as micro-managers start to feed off the same capability for their own managerial decisions (hiring, firing, navigation, contract formation and so forth). This further reduces economic friction, and so on.
7. Enterprise artisan data science is severely challenged, because in the limit, trade is sufficient. In the vast majority of cases, economics dictates that algorithms summon the humans, if necessary, not the other way around.

Like a micro-manager meandering through the prediction web, I find I bump into several somewhat unrelated fields in an effort to refine this hypothesis.

• Errors-in-variables models, optimization, and automated machine learning. In Chapters ?? and ?? I consider the challenge of purely algorithmic management of the production of prediction.
• (Micro) economics. In Chapter ?? I argue that a microprediction network addresses the problem of local knowledge much more effectively than other organizing principles.

- Contest theory and practice. In Chapter ?? I argue that fan out of microprediction tasks is likely to be effective given the theoretical efficacy of contests—not to mention a key role they have played catalyzing the machine learning revolution.
- Control and reinforcement learning. Chapter ?? examines the interplay between microprediction and well worn, effective techniques in control theory and reinforcement learning.
- Privacy preserving computation. Chapter ?? considers the coming wave of federated and outsourced analytics, and some reasons why microprediction capability can move through the seemingly impermeable membranes separating private firms.

I hope you find them as interesting as I do.

2
Oracles

This chapter refines the notion of a microprediction oracle, and infers some minimal properties. An oracle is:

> An apparatus that you can reward for a frequently repeated prediction task, on an ongoing basis, with the expectation that the accuracy will be *eventually hard to beat* on a dollar for dollar basis.
> —Microprediction oracle definition

The physical manifestation is not our immediate concern—though we imagine a formula in an Excel spreadsheet, a Python function, a callback in a web application, a programming interface, or some use of an event processing or messaging system.

2.1 Eventually Hard to Beat

In computer science and mathematics the term oracle is used differently. It refers to a mysterious source of knowledge with *perfect* prescience.

A microprediction oracle as defined is imperfect but nonetheless powerful. The challenge for us is the construction of a forecasting function (to pick one incarnation) that is *as accurate as it can be sooner or later*. The reader is welcome to construct other criteria for the quality of predictive tools—but it is something of a philosophical quagmire.

I prefer to turn the nebulous nature of accuracy on its head with this definition and then reason as to what must be performed by anything trying to meet it. And I will try to exhibit a construction—by which I mean a micro-manager—that according to less than rigorous logic, only fails by a cost factor of two.

The oracle definition expresses an ambition, for micro-managers, that is a cousin to asymptotic efficiency. In statistics we are occasionally able to devise a procedure for estimating a quantity in such a way that *eventually* (as the amount of data tends to infinity) no other method can do it better (i.e. determine a parameter's value with less error).

Now yes, it *is* possible that your data is so well understood that you happen to know a deterministic function that is sufficient to meet the oracle definition. For instance, your data might be generated by a random walk with noise added. There may, by construction, be no possible exogenous data that could possibly help. If you further know the parameters of that process, then the celebrated Kalman filter is certainly *hard to beat*.

However for the vast majority of real-world data we will never know the true generating process. Our search for models that predict best will be never-ending, and our search for data that helps to predict it also. New data is being created all the time.

Yet we wish to design some apparatus that can nonetheless answer our sequence of forecasting questions in the best way possible, eventually. In a practical sense we must bolster the claim, using engineering, that our candidate oracle is hard to beat, now or later.

But how can *any* analytic capability be said to be hard to beat? The marketing material from an AI vendor you are reading presently assures you that their stuff is the bomb—though compared to what? And for how long? Could the errors contain residual noise?

In computer science there is a maxim: write the test first. The criterion *eventually hard to beat for the same or lessor cost* is a high bar and I assert that this mandates active open competition behind the wanna-be oracle—for how else could we hope to know that the performance is the best we can hope for, pound for pound?

2.2 Call a Friend

Now a small logical jump.

I assert that in addition to whatever else it does, a candidate oracle should allow anyone else to contribute predictions that might beat the status-quo (at least for a subset of those questions demanded); that it should seek contributions from competitors and strangers, appropriately assessing their contribution; that it must include some way of combining the results of all who participate (even if that means simply choosing one); and that this should occur in real time.

Chapter **??** will introduce some categories of micro-managers that are modeled on contests, exchanges or cost-aware regression. Any of these, as well as others the reader might devise, might satisfy these requirements.

As an aside, Chapter ?? also explores why "occur in real time" can have a loose interpretation, but Chapter ?? carries some warnings about the practical relevance (or lack thereof) of contest-like mechanisms which operate on fixed, increasingly stale data sets. The judging of modeling contributions using "prediction of the past" has a checkered history.

As a further aside, the oracle might be sneaky, even if it decides to challenge every possible algorithm it finds in real-time. It need not forward every single question that the user of the oracle asks. It might be sparing in its communication.

It might communicate with the world in ways that seem obtuse, in order to preserve privacy.

Its efforts to improve itself might be incremental or episodic in nature, as can its rewards be, and we will consider various possibilities in due course.

However, I do not wish those possibilities to obscure the larger point, and that is that the definition of an oracle does imply some communication with external sources of data and external sources of statistical sneakiness (i.e. modeling approaches) that are not known inside the closed system that might constitute the oracle's code base, or the oracle's data stores, or the oracle's complex event processing system, et cetera.

Table 2.1 is intended to provide a glance at the fanout in application taxonomy suggested merely by examples which—would you believe—have already been the subject of competitive prediction. I have taken some data science contest examples and placed them in respective categories, or rather sub-sub-categories.[1]

That's the best I can do in this format, but in reading Table 2.1 one has to appreciate that in addition to "image" being a subcategory of "recognition" we also have subcategories for motion, audio, text, EEG.

Because we cannot know the best way to combine models, and we cannot know the best way to manage in automated fashion a collection of models, we are drawn to something like Figure 3.1 which is intended to represent a tower of micro-managers all trying to out-predict someone and, to better their chances, enlisting the expertise of other micro-managers who have access to additional techniques and data.

If the top-level oracle is reactive (operates like a function responding to your prodding) then the more time given to respond to the question, and the lower the overhead for running a micro-manager (i.e. fanout, management and so on) the deeper the calculation tree. The more micro-managers are involved, the greater the opportunity for specialization—not to mention broad data search.

Table 2.1
Fanout in taxonomy of categories of applications for crowd-sourced micro-prediction. For each major category we list only one sub-category. For each sub-category I list only one sub-sub-subcategory.

Category	Example Sub-cat.	Example Sub-sub cat.
Recognition	Image	Facial
Search	Travel	Personalization
Recommendation	Ad-tech	Click-throughs
Government	Open cities	Flight status
Sales and CRM	Repeat shopping	Visitation
Internet of things	Homes	Usage
Environment	Air	Pollutants
Transport	Driving	Distracted driver
Manufacturing	Industrial control	Predictive maintenance
Agriculture	Juice	Orange juice
Finance	Investment banking	Commercial loans
Energy	Power	Wind
Medicine	Inventory	Hospital stays

This diagram is intended to also convey some of the thorny questions that arise for micro-managers trying to survive in this game.

The oracle property is evidently not a quality of a single micro-manager. Rather, it emerges over time when a complex wiring of inter-related micro-managers is driven by greed.

But there must be something wrong with this proof, given the lack of existence of a microprediction web used universally across all industries! I posit:

1. An inadequate supply of rewarded microprediction tasks.
2. Trade frictions

My hope is that advertising the possibility of a prediction web might help with the first issue. Let's turn again to the second.

3.6 Overhead

It is fair to say that our pseudo-oracle *encourages* a jumbled collection of nested and recurrent calls to micro-managers, including other pseudo-oracles (that precedes some eventual higher level recombination—and thereby an answer returned to the user). In conjunction with Chapter ??, this line of reasoning hopes to suggest that competitive tension is *necessary*.

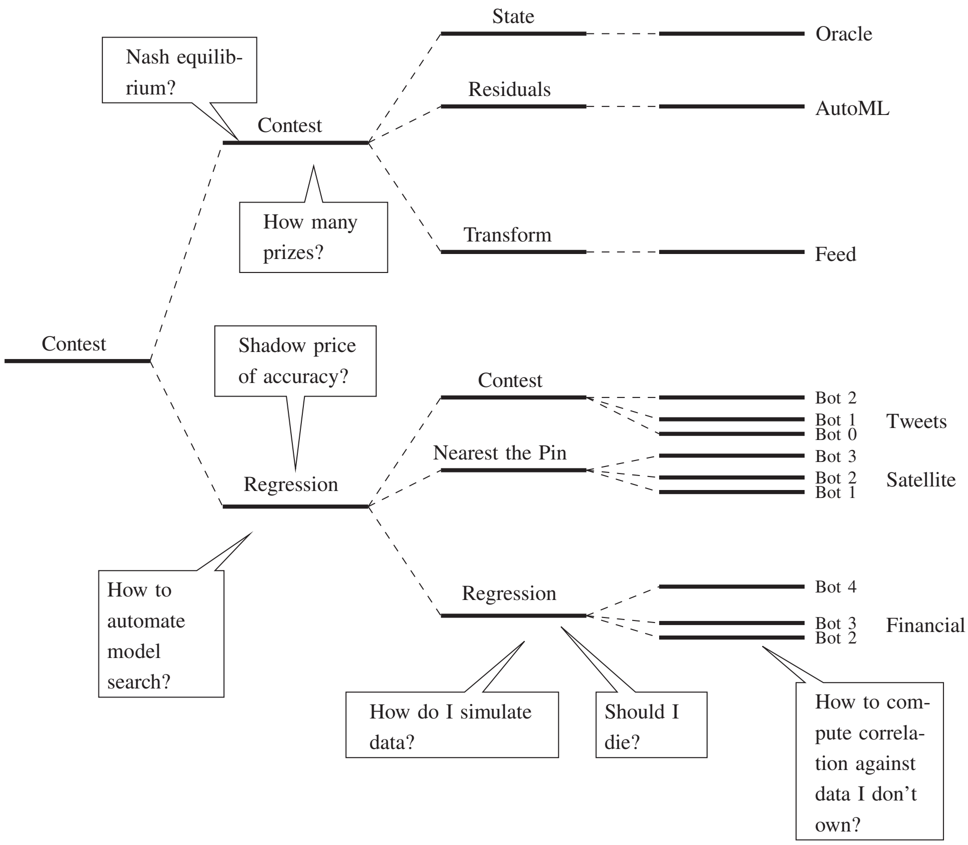

Figure 3.1
Part of a microprediction web. An oracle answers prediction questions by fanning them out to micro-managers in real time, assessing longitudinal performance, and preparing an ensemble response. Algorithms can do the same, leading to a supply chain. Some managerial issues are suggested by questions, and discussed in later chapters.

Is it sufficient? One might cautiously say yes, the oracle definition appears to be met. For if it was easy to beat the output of this collective calculation, say with some new variety of neural network, then it should be relatively easy to endow said challenger with sufficient navigation ability (and economic common sense) that it finds its way across the network and addresses the problem. It is fair to say that our pseudo-oracle *encourages* a jumbled collection of nested and recurrent calls to micro-managers, including other pseudo-oracles (that precedes some eventual higher level recombination—and thereby an answer returned to the user). In conjunction with Chapter ??.

-1
0
+1

3.6 Example

Confident in the march of technology, let's make our pseudo-oracle more concrete. This example is intended to illustrate that an "edge" micro-manager (aspiring oracle, or pseudo-oracle) that faces application doesn't need to be perfect.

3.6.1 When will the School Bus Arrive?

What a fortunate life you live. Every day a yellow school bus pulls up to the top of your driveway and delivers your offspring. Your only task is to rush out—typically in the middle of a meeting—and meet them.

The application we have in mind is purely passive. You do not do anything, beyond wearing your watch. However your watch will provide you with a two minute warning. Your driveway is rather long, we shall assume.

Perhaps—and now we're getting fancy, you might input some indication of your cost function—which is to say how annoyed the school bus driver will be if you consistently leave them waiting.

The application must be hungry for predictive intelligence. It will face competition from other applications like it. In the absence of a prediction web, the development of a state of the art predictive model might chew up months.

That model would be constantly revised. Data scientists might struggle with cleaning, extracting and use of data effectively—data drawn from a wide variety of sources with different formats.

In order to remain popular in the *eventually* competitive bus-arrival genre *forever*, the application *eventually* needs to know all sorts of things. It needs data scraped from local sources. The bus is early when there is no band practice, and late during certain special events, and later yet when Mrs Meldrum, your neighbor down the street, is asleep in her house (the driver has had to wake her once or twice—her absence from social media is a status clue).

The application logic, as distinct from the prediction logic, is pretty trivial. Your watch notices two events. The first event is your arrival near the top of your driveway. The second is your walking away. Typically the second event occurs several minutes after the first—or at least it used to before the invention of the prediction web.

But now life is better because your watch can ping our pseudo-oracle, multiple times as needed, and receive good updated predictions of the estimated time of arrival. We imagine this is a cloud web service in this particular example (it wakes when a question arrives, does something, responds to you, and then goes back to sleep.)

The second event, your walking away from your letterbox, triggers a different type of message from your watch to the aspiring oracle. It sends an approximate ground truth, and while not absolutely necessary in every application, this certainly simplifies things.

The running cost of this edge micro-manager is measured in hundredths of a cent per month. Can it deliver that much value to you? We are about to find out because today, your watch held back on buzzing you for two minutes longer than usual. That's two whole minutes of sunshine you won't get, in your over-optimized life.

Still, it did give you time to finish that email, before walking out to meet your children.

3.6.2 Oracle Implementation

What does the pseudo-oracle do? As suggested by our reasoning, the oracle is a relay station that allows external people and algorithms to prove that they can predict the arrival time accurately. Put another way, it serves as an ongoing, real-time contest, as suggested by Figure ??, and in this example it will do very little else.

At the moment the oracle receives your question ("when will the bus arrive?") it relays the question to a subset of the algorithms that have registered their interest. It will use its children to predict where yours are.

The watch app has conveyed, as part of the question, that it expects an answer within five seconds, so the oracle knows it had better establish a deadline for responses from children. That will allow it time to combine the results, and relay them to the watch.

The oracle, or aspiring oracle we must remind ourselves, will also help initiate a relationship with other algorithms. We shall suppose it is implemented entirely reactively.

That is to say that the micro-managing pseudo-oracle answers the question "when's the bus coming?" but it also answers other types of questions that help other algorithms decide if they want to participate. For instance, the oracle might respond to questions like "what's the prizemoney?" with "not much, suck it up".

A relatively simple, well understood model for bus arrivals is used by the oracle to generate lower and upper bounds. This model assumes that the best estimate for the arrival time of the bus is an average of the last ten arrival times. It's not a great model, but regulators can understand it instantly, and it doesn't take long to document.

3.6.3 A Modified Median Filter

Upon receiving answers from the children the oracle must combine them and form a response for the watch. The first child's recency-weighted score is suggestively denoted V_1, the second V_2 and so forth. The oracle ranks the children by their (inaccuracy) scores from lowest to highest. The oracle then returns the median of the answers to the current question provided by the five children with lowest scores.

The exception is when that median differs from the approved model's answer by more than three minutes, whereupon the oracle adjusts the estimate closer to the approved model's answer until this is no longer the case.

A few minutes later the oracle receives the truth message. The bus actually arrived at 3.27pm. The first child predicted 3:28:10. The error in the estimate given by the first child is 70 seconds. The squared error is 4900. The aspiring oracle updates the child's score

$$\overbrace{V_1(new)}^{\textit{updated performance}} = 0.99 * \overbrace{V_1(old)}^{\textit{previous}} + \underbrace{0.01}_{\textit{speed}} * \overbrace{4900}^{\textit{score}}$$

and does the same for all children. In the event that a child fails to respond, or provides a badly formatted answer, it is put in the dog-house (just as you will be if you are consistently late). The child's score is reset:

$$\overbrace{V_1(new)}^{\textit{updated}} = \underbrace{360000}_{\textit{reset inaccuracy}}$$

This amounts to a presumed standard error of ten minutes, which we assume is pretty poor.

3.6.4 Payouts

I assume that periodically the oracle pays out accumulated prizes (received from the watch application) to children, according to their accuracy measured over some epoch (say a month). The most accurate model receives *half the total reward*. Less accurate contributors receive smaller prizes of $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, $\frac{1}{16}$, based on both the accuracy and originality of their contribution.

A simple way to define originality is by establishing an epoch-based priority scheme modeled after the patent system. Contributions are disqualified if they are consistently close to a contribution with greater vintage.

3.6 Analysis

This completes the description of a very elementary attempt at an oracle. It may seem unlikely to meet our definition, but let's fast-forward several years. Perhaps by this time the arrival estimates are extremely accurate because everyone in your neighbourhood is using it, thereby generating more data and rewards.

Perhaps by this time the small rewards have prompted someone to put a tracking device on the bus itself, thereby driving the prediction error to essentially zero. There is no such thing as cheating, in this particular contest.

3.6.1 Accomplishments

The supply chain we have encouraged:

1. Permits granular contributions,
2. and thus specialization,
3. and also cross subsidy (sharing of data and features),
4. and possibly new data creation,
5. with zero *mandated* human management or the inevitable overhead.

There are some possible downsides too, beginning with fear of opaque and changing model processes.

3.6.2 Model Risk

The beauty of the microprediction domain is that sometimes the so-called model risk is bounded.

We merely observe that the oracle's predictions are "explainable within three minutes" and with that, the materiality of the model risk has been capped.

Bounding model risk isn't the same as bounding risk, it merely establishes that the risk is no worse, materially, than it is with any other solution that is considered to have acceptable model risk.

(Please don't complain. Some if not most model risk policies are ridiculous due to the Stone-Weierstrass theorem, which can be translated to state that models with allegedly comprehensible risk are dense in the space of all models. What logic would you have me build on top of that quicksand?)

The use of the approved model isn't "free", by the way. If the bus can't start one day, and several models with superior data can discern this, too bad for you. The circuit-breaker will prevent you from being alerted to the full extent of the delay and therefore you may wait quite a while—albeit three minutes less than if you had relied only on an approved model.

So by various metrics there is a cost to using the approved model. However on an amortized basis this cost is much smaller than other possible costs that might be imposed upon us—such as documenting, and re-documenting, and re-documenting the operation of all models that contribute as they constantly improve.

3.6.3 Resistance to Manipulation

The aspiring oracle is designed with simplicity and robustness in mind.

That's not to suggest we shouldn't be wary. A denial attack occurs when a participant clones the same entry many times (with a tiny amount of noise pollution). We will assume that this is defeated by the priority defence—though in practice there are other defenses too.

A Sybil attack, as we might term it, occurs when a nefarious source of intelligence creates three highly accurate prediction algorithms that are quite different yet outperform all others in the contest (the adversary may likely invest more effort than is warranted by the small rewards, in order to pull this off). Over the course of several months these entrants achieve the lowest error, thus taking control of the median.

Then, one day, they conspire to return a bogus answer with the intent of deceiving you. The attack may persist for some number of days until the self-inflicted performance penalty moves the deceiving algorithms down the leaderboard.

This doesn't cause great harm. The existence of an approved model and the bound means that the estimate will still be within three minutes of a half-way plausible guess. Not a catastrophic malfunction, by any means.

3.6.4 Reliability

No algorithm supplying predictions can be a single point of failure—except the approved model which is too dumb to fail. By encouraging diversity we become less reliant on any one algorithm performing when needed. We can fall back to the approved model, if need be.

(Needless to say, there are more sophisticated methods of encouraging diversity than the simple system used in this example.)

To receive *any* payment an algorithm must also be accurate in its own right. So, if the top supplier drops out, the degradation in quality should not be extreme. There is incentive for all algorithms, however original, to improve their accuracy and maintain a spot near the top.

There's a failure mode involving repeated fallback to the approved model—but that's easily detected. The oracle could even use another oracle for more advanced anomaly detection.

3.6.5 Eventually Hard to Beat?

Now for the real exam.

I turn to the question of the eventual efficiency—which is touted as the main selling point of the oracle. We observe that there are many things that might be considered questionable about the design. I have chosen a particular combination of model aggregation, and reward scheme, that is reasonably inoffensive, but not likely to be the best in any sense.

I've noted in generality that there are many ways to reward and combine, and no way to know in advance what the best one will be. Even within the strictures of this narrow example, there are many choices to be made, such as the number of children (5) to include in the median, the speed parameter (0.01), the reset inaccuracy score $(360,000)$, the exponent (2) used to covert error into score, the count-back mechanism for prizes, and the payout fractions $\frac{1}{2}, \frac{1}{4}, \ldots$.

Value won't be injected into the network only by Ph.D.'s developing cutting edge techniques which are published in the top journals. On the contrary, tiny automated middle-people can be authored by anyone.

We should respect opportunism in both the authors and their product. Speaking to the knowledge possessed by the real estate broker whose "knowledge is almost exclusively one of temporary opportunities", Hayek writes:

> It is a curious fact that this sort of knowledge should today be generally regarded with a kind of contempt and that anyone who by such knowledge gains an advantage over somebody better equipped with theoretical or technical knowledge is thought to have acted almost disreputably.

This applies rather obviously to exogenous data. But to focus on model search momentarily, it is apparent that we have tens of thousands of open source repositories to choose from, often containing the latest discoveries. They are mostly, woefully underutilized.

> To gain an advantage from better knowledge of facilities of communication or transport is sometimes regarded as almost dishonest, although it is quite as important that society make use of the best opportunities in this respect as in using the latest scientific discoveries.

Some would say the academic system serves this purpose. There is some reward offered, in the sense that it is easier.

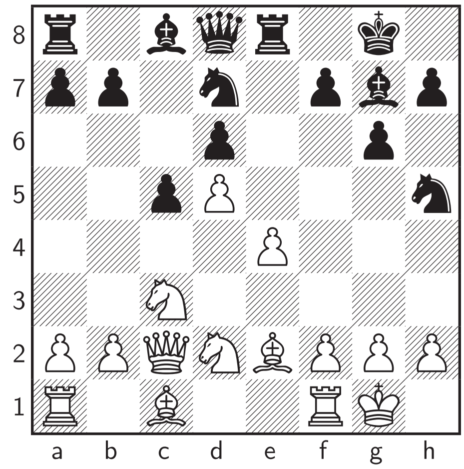

Figure 8.1
Spassky-Fischer World Championship Game 3 1972. Position after Fisher's anti-positional move 11..Nh5!?

I turn to the question of the eventual efficiency—which is touted as the main selling point of the oracle. We observe that there are many things that might be considered questionable about the design. I have chosen a particular combination of model aggregation, and reward scheme, that is reasonably inoffensive, but not likely to be the best in any sense. To receive *any* payment an algorithm must also be accurate in its own right. So, if the top supplier drops out, the degradation in quality should not be extreme. There is incentive for all algorithms, however original, to improve their accuracy and maintain a spot near the top. There is incentive for all algorithms, however original, to improve.

I suggest that an oracle using the simplest derivative pattern might not always be best, and that, therefore, moving beyond chess, we should judge conditional estimates $V(state|action)$ against *future oracle answers*, not only the next value estimate $V(state)$, unless there is some strong reason to have great faith in $V(state)$.[2]

That's not controversial and in fact I'm taking[3] a leaf out of the temporal difference learning literature.[4] Here is an example of a truth that might be generated ex-post. We ask algorithms to aim at a target that is a weighted[5] combination of future position evaluations:

$$target = (1 - \lambda) \sum_{k=1}^{\infty} \lambda^{k-1} V(k \text{ moves ahead}) \tag{8.1}$$

for some choice[6] of λ between zero and one.[7] For example with $\lambda = \frac{1}{2}$ the forward [8] target would read

$$\begin{aligned} target = &\frac{1}{2} V(1 \textit{ move ahead}) + \\ &+ \frac{1}{4} V(2 \textit{ moves ahead}) + \\ &+ \frac{1}{8} V(3 \textit{ moves ahead}) + \\ &+ \frac{1}{16} V(4 \textit{ moves ahead}) + \\ &+ \dots \end{aligned}$$

whereas with $\lambda = 0.95$ we weight future move evaluations more heavily:

$$\begin{aligned} target = &\frac{95}{1000} V(1 \textit{ move ahead}) + \\ &+ \frac{90}{1000} V(2 \textit{ moves ahead}) \\ &+ \dots \\ &+ \frac{60}{1000} V(10 \textit{ moves ahead}) \end{aligned}$$

A micro-manager might[9] choose $\lambda = 0.25$ and another manager[10] might choose $\lambda = 0.95$ say, with only the latter rewarding[11] those who judged the game to be going in Fischer's favour immediately after 11..*Nh5*!?.[12]

__-1
__0
__+1

Afterword

It is hard, you will agree,[13] for any considering human to avoid forming a mental picture of the way that quantitative techniques will influence commerce, and society, going forward.[14] Let's face it, *all future visions* of *just about anything* place mathematically driven automation (by various names) in a central role.[15]

(Even the fictional President Josiah Bartlett, of *The West Wing*, told us that the 21st Century would be the century of statistics.)

Evidently my own vision includes a new type of utility. The task of catalyzing a prediction web is better addressed by code, not prose, and I really must get back to that. But I hope you take some time to consider the potential for competitive prediction in the "small".

As with computing prophesies of yesteryear that only saw enterprise use,[16] I suspect we are mostly blind to what the equivalent of an in-house data science team will eventually be used for, once shrunk down to the size of a thimble. I'm sure I'll look back on these pages in a few years and kick myself for missing something obvious.

I'm reminded too of how difficult it is to overcome our mental inertia, especially on the matters of cost - the central obsession of this book. Many years ago my undergraduate professor, filling out a form to acquire me a university computer account, listed "email" as the only justification. I was shocked at the time as computers were expensive. It seemed unlikely that an administrator would allocate precious computation and bandwidth to someone whose only stated use was so frivolous.

Needless to say I got that computer account, and how quaint that guilt seems now. In the same way, I think we will use a network of micro-predicting algorithms and data in ways that initially seem fatuous. Later, the exact same uses will be considered essential. The idea of not being able to map the near future, of everything large and small, will seem as antiquated as a car that cannot see the road ahead.

-1
0
+1

Admittedly, we won't all use microprediction at the lowest level of implementation detail, but that's okay. We aren't all experts in the TCP/IP internet protocols, either. Whatever your prefered level of abstraction is, I hope you are interested in helping to create an incredibly inexpensive alternative to the "data science project". A universal source of intelligence, despite the limitations I've stated, could be the next great utility.

To emphasize one last time, I have concerned myself with the microprediction domain only, for reasons that I hope are now clear. It is not *the* future of AI. It is not *the* future of statistics. It is not the future of general artificial intelligence - merely half the things branded AI. As I write these words, the world is experiencing a dreadful pandemic. It needs thoughtful inferential statistics to interpret a dire situation, make medium term forecasts, and approve vaccines.

This kind of statistics will never be replaced by a microprediction network. But a substrate where algorithms travel can help in surprising ways - as with the sourcing of surrogate models for disease spread, or crowd-sourced approximations for long-running molecular simulations.

And *in addition* to medium term prediction, or the things we usually associate with the word prediction, the world also needs inexpensive, accessible but accurate microprediction to drive operational efficiency in every industry. This is well within our collective ability. The prediction web is ambitious in some ways, but rather mundane in others. It almost feels like busywork.

Notes

Chapter 1

1. Silver et al. (2021)

Chapter 2

1. In the taxonomy I'm drawing upon numerous sources including some mentioned in Chapter ??. Kaggle provides more data than most.Kaggle Inc. (2014)

2. For an example of how microstructure complicates econometric estimation see Hansen and Lunde (2006) and O'Hara (2015)

3. "Nibblefish" was once a candidate name for the prediction web project, with rather mixed focus group results. I urge the reader to consult the Scottish Health Protection agency report on risks associated with *Garra rufa* fish pedicures before engaging in that activity.Agency

4. Definition of redlining is from Wikipedia.

5. Credit redlining is discussed in Cohen-Cole (2011). Groceries in Eisenhauer (2001). See Lang and Nakamura (1993) for further discussion.

6. The New York City Council passed legislation on December 11, 2017 according to an article published by the UCLA Richardson (Dec 2017)

7. Turek (2017). Also emphasized is the enabling of human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

8. See Forrester and Keane (2009) for a survey of surrogate methods for optimization, though plenty has occurred since.

9. In the taxonomy I'm drawing upon numerous sources including some mentioned in Chapter ??. Kaggle provides more data than most.Kaggle Inc. (2014)

10. Wikipedia was also directly inspired by Hayek's essay.Tucker (Mar 2017)

11. The mentioned course is offered by Stanford University and Coursera

12. See Freider et al. for an example of a pipeline tool.

13. Examples of meta-data tooling include the DataHub and Amundsen packages.

14. Other competitive sites more focused on evaluation of participants, (rather than solving industry problems) include Hackerearth, Interviewbit, Hackerrank, CodeGround, Codeforces, Sphere Online Judge, UVA online judge and Code Chef. These collectively add another million or so users. Numbers have increased substantially since compilation or the date of sources used. I was not able to find estimates for some notables such as WorldQuant, Numerai or Quantiacs, and apologize for any omissions.

15. The description of the Common Task Framework echos Donoho Donoho (2015).

16. Kaggle reported a 340 percent accuracy improvement, which sounds good, even if I can't help you parse that number.

Bibliography

GE Flight Quest, 2013. URL http://www.gequest.com/c/flight/details/winners.

Michael Abramowicz. The Hidden Beauty of the Quadratic Market Scoring Rule: a Uniform Liquidity Market Maker, with Variations. *The Journal of Prediction Markets*, 1(2):111–125, 2007. URL http://www.ingentaconnect.com/content/ubpl/jpm/2007/00000001/00000002/art00002.

Philip Adler, Casey Falk, Sorelle A. Friedler, Tionney Nix, Gabriel Rybeck, Carlos Scheidegger, Brandon Smith, and Suresh Venkatasubramanian. Auditing black-box models for indirect influence, 2017. ISSN 02193116.

Health Protection Agency. Guidance on the management of the public health risks from fish pedicures. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/322420/Fish_Spa_guidance.pdf.

Ayman Alharbi, Haneen Zamzami, and Eman Samkri. Survey on homomorphic encryption and address of new trend. *International Journal of Advanced Computer Science and Applications*, 11(7), 2020. ISSN 21565570. doi: 10.14569/IJACSA.2020.0110774.

Moez Ali and Antoni Baum. PyCaret Python Package, 2020. URL https://github.com/pycaret/pycaret.

Melissa Assel, Daniel D. Sjoberg, and Andrew J. Vickers. The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagnostic and Prognostic Research*, 2017. doi: 10.1186/s41512-017-0020-3.

Robert Axelrod. Effective Choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, 24(1), 1980. ISSN 15528766. doi: 10.1177/002200278002400101.

Ken Baron and Jeffrey Lange. *Parimutuel Applications in Finance*. 2007. doi: 10.1057/9780230627505.

Pauline Bennet, Carola Doerr, Antoine Moreau, Jeremy Rapin, Fabien Teytaud, and Olivier Teytaud. Nevergrad. *ACM SIGEVOlution*, 14(1), 2021. doi: 10.1145/3460310.3460312.

James Bennett and Stan Lanning. The Netflix Prize. *KDD Cup and Workshop*, pages 3–6, 2007. ISSN 1554351X. doi: 10.1145/1562764.1562769.

G. E. P. Box and D. R. Cox. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 1964. doi: 10.1111/j.2517-6161.1964.tb00553.x.

L Breiman. Statistical Modeling: The Two Cultures. *Statistical Science*, 2001. ISSN 08834237. doi: 10.2307/2676681.

Mark Broadie and Soonmin Ko. A simulation model to analyze the impact of distance and direction on golf scores. In *Proceedings - Winter Simulation Conference*, 2009. doi: 10.1109/WSC.2009.5429280.

Karl Broman. Data Science is Statistics, Apr 2013. URL https://kbroman.wordpress.com/2013/04/05/data-science-is-statistics/.

Jason Byrne. Persistent Data in Lambda. It Works!, Dec 2018. URL https://engineering.flosports.tv/persistent-data-in-lambda-it-works-ca0c1b25879e.

Coralia Cartis, Jan Fiala, Benjamin Marteau, and Lindon Roberts. Improving the flexibility and robustness of model-based derivative-free optimization solvers. *ACM Transactions on Mathematical Software*, 45 (3), 2019. ISSN 15577295. doi: 10.1145/3338517.

H Coase, R. Nature of the firm. *Economica*, 1937.

Ethan Cohen-Cole. Credit Card Redlining. *The Review of Economics and Statistics*, 93(2):700–713, 2011. ISSN 00346535. doi: 10.1162/REST{_}a{_}00052</p>. URL http://dx.doi.org/10.1162/REST_a_00052%5Cnhttp://www.mitpressjournals.org/doi/abs/10.1162/REST_a_00052.

Brian Conlon. Totalizator History, 1997. URL http://members.ozemail.com.au/~bconlon.

Rusty Conover. Predicting New York State's Electricity, Jan 2021. URL https://rusty.today/posts/predicting-ny-electricity-with-julia-and-tensorflowjs.

Peter Cotton. The future of forecasting, according to the experts. URL https://www.microprediction.com/blog/future.

Peter Cotton. Self Organizing Supply Chains for Micro-Prediction: Present and Future uses of the ROAR Protocol, 2019a. ISSN 23318422.

Peter Cotton. How to Respond to an RFQ, 2019b. URL https://github.com/microprediction/winning/blob/main/docs/How_to_respond_to_an_RFQ.pdf.

Peter Cotton. Why liberals should be terrified right now - how a trump surge can sneak past the election models, 2020. URL https://www.microprediction.com/blog/election.

Peter Cotton. Inferring Relative Ability from Winning Probability in Multientrant Contests. *SIAM Journal on Financial Mathematics*, 12, 2021a. doi: 10.1137/19M1276261.

Peter Cotton. HumpDay Optimization Package, 2021b. URL https://github.com/microprediction/humpday.

Peter Cotton. The Only Prediction Function You'll Ever Need?, Aug 2021. URL https://www.microprediction.com/blog/forever.

Peter Cotton. The Lottery Paradox, Dec 2020a. URL https://www.microprediction.com/blog/lottery.

Peter Cotton. Popular Python Time-Series Packages, Dec 2020b. URL https://www.microprediction.com/blog/popular-timeseries-packages.

Peter Cotton. Time Machines Python Package, Dec 2021. URL https://github.com/microprediction/timemachines.

Peter Cotton. Is Facebook's Prophet the Time-Series Messiah, or Just a Very Naughty Boy?, Feb 2021. URL https://www.microprediction.com/blog/prophet.

Peter Cotton. Where Will a Badminton Player Move to Next, and How Should we Adjudicate Predictions of the Same?, July 2020a. URL https://www.microprediction.com/blog/badminton.

Peter Cotton. Microprediction.Org, July 2020b. URL https://www.microprediction.org.

Ronald Cramer, Ivan Bjerre Damgård, and Jesper Buus Nielsen. *Secure multiparty computation and secret sharing*. 2015. doi: 10.1017/CBO9781107337756.

Robert H. Crites and Andrew G. Barto. Elevator Group Control Using Multiple Reinforcement Learning Agents. *Machine Learning*, 12(4), 1998. ISSN 08856125. doi: 10.1023/a:1007518724497.

Pedro Domingos. *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World.* Basic Books, 2015.

David Donoho. 50 Years of Data Science. *R Software*, page 41, 2015.

DTCC. DTCC Swap Data Repository Realtime Dashboard, 2019. URL https://rtdata.dtcc.com/gtr/.

Jiexin Duan, Xingye Qiao, and Guang Cheng. Statistical guarantees of distributed nearest neighbor classification. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 229–240. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/022e0ee5162c13d9a7bb3bd00fb032ce-Paper.pdf.

Werner Ehm, Tilmann Gneiting, Alexander Jordan, and Fabian Krüger. Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 2016. ISSN 14679868. doi: 10.1111/rssb.12154.

Elizabeth Eisenhauer. In poor health: Supermarket redlining and urban nutrition. *GeoJournal*, 53(2):125–133, 2001. ISSN 03432521. doi: 10.1023/A:1015772503007.

Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505*, 2020.

Arno Candel et al. *H20 AutoMl*, 2015. URL https://github.com/h2oai.

Holder Hurtgen et al. Rethinking AI talent strategy as automated machine learning comes of age, 2020.

James Manyika et al. Big data: The next frontier for innovation, competition and productivity, May 2011. URL https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation.

Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter. Efficient and Robust Automated Machine Learning. *Advances in Neural Information Processing Systems 28*, pages 2944–2952, 2015. ISSN 10495258. URL http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf.

Alexander I.J. Forrester and Andy J. Keane. Recent advances in surrogate-based optimization, 2009. ISSN 03760421.

Colin Frayn. The chess brain project. URL http://chessbrain.net/.

Elias Freider, Erick Bernhardsson, Arash Rouhani, and Dave Buchfuhrer. Luigi package. URL https://luigi.readthedocs.io/en/stable/.

Wayne A. Fuller. *Measurement Error Models*. Wiley, 1987.

Craig Gentry. *A fully homomorphic encryption scheme*. PhD thesis, 2009.

Pieter Gijsbers and Joaquin Vanschoren. GAMA: Genetic Automated Machine learning Assistant. *Journal of Open Source Software*, 4(33), 2019. ISSN 2475-9066. doi: 10.21105/joss.01132.

Robby Goetschalckx, Kurt Driessens, and Scott Sanner. Cost-sensitive parsimonious linear regression. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2008. doi: 10.1109/ICDM.2008.76.

Peter R. Hansen and Asger Lunde. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24 (2), 2006. ISSN 07350015. doi: 10.1198/073500106000000071.

F Hayek. The Use of Knowledge in Society FA Hayek. *American Economic Review*, 1945. ISSN 00028282. doi: 10.1257/aer.98.5.i.

Bengt Holmstrom and Paul Milgrom. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica*, 1987. ISSN 00129682. doi: 10.2307/1913238.

TJ Holowaychuk. AWS Lambda lifecycle and in-memory caching, 2017. URL https://medium.com/@tjholowaychuk/aws-lambda-lifecycle-and-in-memory-caching-c9cd0844e072.

Mohammad Zakir Hossain. The use of Box-Cox transformation technique in economic and statistical analyses. *Journal of Emerging Trends in Economics and Management Sciences*, 2(1), 2011.

Joshua Introne, Robert Laubacher, Gary Olson, and Thomas Malone. The Climate CoLab: Large scale model-based collaborative planning. In *Proceedings of the 2011 International Conference on Collaboration Technologies and Systems, CTS 2011*, 2011. doi: 10.1109/CTS.2011.5928663.

Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. Classification with costly features using deep reinforcement learning. In *ALA 2018 - Adaptive Learning Agents - Workshop at the Federated AI Meeting 2018*, 2020.

Haifeng Jin, Qingquan Song, and Xia Hu. Auto-keras: An efficient neural architecture search system. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019. doi: 10.1145/3292500.3330648.

Matthew Joseph, Aaron Roth, Jonathan Ullman, and Bo Waggoner. Local differential privacy for evolving data. *Journal of Privacy and Confidentiality*, 10(1), 2020. ISSN 25758527. doi: 10.29012/jpc.718.

Kaggle Inc. The Home of Data Science, 2014. URL http://www.kaggle.com.

Edward H. Kaplan and Candler Rich. Decomposing Pythagoras. *Journal of Quantitative Analysis in Sports*, 2017. ISSN 15590410. doi: 10.1515/jqas-2017-0055.

Jonathan P Kastellec. The Statistical Analysis of Judicial Decisions and Legal Rules with Classification Trees. *Journal of Empirical Legal Studies*, 7(2):202–230, 2010. ISSN 1740-1461. doi: 10.1111/j.1740-1461.2010.01176.x.

Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 2009. ISSN 15324435.

Brent Komer, James Bergstra, and Chris Eliasmith. Hyperopt-Sklearn: Automatic Hyperparameter Configuration for Scikit-Learn. In *Proceedings of the 13th Python in Science Conference*, 2014. doi: 10.25080/majora-14bd3278-006.

William W. Lang and Leonard I. Nakamura. A Model of Redlining. *Journal of Urban Economics*, 33(2):223–234, 1993. ISSN 00941190. doi: 10.1006/juec.1993.1014. URL http://linkinghub.elsevier.com/retrieve/pii/S0094119083710144.

Tamir Levy and Joseph Yagil. Air pollution and stock returns in the US. *Journal of Economic Psychology*, 2011. ISSN 01674870. doi: 10.1016/j.joep.2011.01.004.

Xueqing Liu and Chi Wang. An Empirical Study on Hyperparameter Optimization for Fine-Tuning Pre-trained Language Models. 2021. doi: 10.18653/v1/2021.acl-long.178.

Spyros Makridakis, Chris Fry, Fotios Petropoulos, and Evangelos Spiliotis. The future of forecasting competitions: Design attributes and principles. URL https://arxiv.org/abs/2102.04879.

Spyros Makridakis, Michele Hibon, and Claus Moser. Accuracy of Forecasting: An Empirical Investigation. *Journal of the Royal Statistical Society. Series A (General)*, 142(2), 1979. ISSN 00359238. doi: 10.2307/2345077.

Thomas W Malone. How Human-Computer 'Superminds' Are Redefining the Future of Work. *MIT Sloan Management Review*, 2018.

Ruben Martinez-Cantin. BayesOpt: A Bayesian optimization library for nonlinear optimization, experimental design and bandits. *Journal of Machine Learning Research*, 15, 2015. ISSN 15337928.

McKinsey & Company. Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, (June):156, 2011. ISSN 14712970. doi: 10.1080/01443610903114527.

Dara D. Mendez, Vijaya K. Hogan, and Jennifer Culhane. Institutional racism and pregnancy health: Using Home Mortgage Disclosure Act data to develop an index for mortgage discrimination at the community level. *Public Health Reports*, 126(SUPPL. 3), 2011. ISSN 14682877. doi: 10.1177/00333549111260s315.

Steven J Miller. The pythagorean won-loss formula and hockey: A statistical justification for using the classic baseball formula as an evaluative tool in hockey. *arXiv preprint arXiv:1208.1725*, 2012.

Steven D. Moffitt and William T. Ziemba. A method for winning at lotteries, 2018. ISSN 23318422.

Maureen O'Hara. High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270, 2015. ISSN 0304405X. doi: 10.1016/j.jfineco.2015.01.003.

Randal S. Olson, Nathan Bartley, Ryan J. Urbanowicz, and Jason H. Moore. Evaluation of a tree-based pipeline optimization tool for automating data science. In *GECCO 2016 - Proceedings of the 2016 Genetic and Evolutionary Computation Conference*, 2016. doi: 10.1145/2908812.2908918.

M I T Opencourseware. Review of the Asymptotics of Extremum Estimators. *Time*, 2008.

Tabitha C. Peck, Sofia Seinfeld, Salvatore M. Aglioti, and Mel Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, 22(3):779–787, 2013. ISSN 10902376. doi: 10.1016/j.concog.2013.04.016.

Devin G. Pope and Maurice E. Schweitzer. Is tiger woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review*, 101(1), 2011. ISSN 00028282. doi: 10.1257/aer.101.1.129.

Tommaso Proietti and Helmut Lütkepohl. Does the Box-Cox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting*, 29(1), 2013. ISSN 01692070. doi: 10.1016/j.ijforecast.2012.06.001.

Amnon Rapoport, Darryl A. Seale, and Andrew M. Colman. Is tit-for-tat the answer? on the conclusions drawn from axelrod's tournaments. *PLoS ONE*, 10(7), 2015. ISSN 19326203. doi: 10.1371/journal.pone.0134128.

Forrester Research. Data Science Marketplaces Unlock Advanced Analytics for the Masses, Jul 2016.

Rashida Richardson. New York City takes on Algorithmic Discrimination, Dec 2017. URL https://www.aclu.org/blog/privacy-technology/surveillance-technologies/new-york-city-takes-algorithmic-discrimination.

R. M. Sakia. The Box-Cox Transformation Technique: A Review. *The Statistician*, 41(2), 1992. ISSN 00390526. doi: 10.2307/2348250.

Leonard J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 1971. ISSN 1537274X. doi: 10.1080/01621459.1971.10482346.

Alan Shipnuck and Rick Reilly. Sports illustrated: The crack-up, 2006.

David Silver, Satinder Singh, Doina Precup, and Richard S. Sutton. Reward is enough, 2021. ISSN 00043702.

Michael Stöckl and Martin Lames. Modeling Constraints in Putting : The ISOPAR Method. *International Journal of Computer Science in Sport*, 10 (1992), 2011.

S J Stolfo. KDD cup 1999 dataset. *UCI KDD repository. http://kdd.ics.uci.edu*, page 0, 1999.

Qichun Tang. Ultraopt : Distributed asynchronous hyperparameter optimization better than hyperopt, 2021. URL https://doi.org/10.5281/zenodo.4430148.

Janek Thomas, Stefan Coors, and Bernd Bischl. Automatic gradient boosting. In *International Workshop on Automatic Machine Learning at ICML*, 2018.

Timgates42. scikit-optimize, 2020.

Jeffrey A. Tucker. Wikipedia Is the Wonder of the World that Wasn't Supposed to Work, Mar 2017. URL https://fee.org/articles/wikipedia-is-the-wonder-of-the-world-that-wasn-t-supposed-to-work/.

Matt Turek. Explainable Artificial Intelligence, 2017. URL https://www.darpa.mil/program/explainable-artificial-intelligence.

Tyler J. VanderWeele. On a Square-Root Transformation of the Odds Ratio for a Common Outcome, 2017. ISSN 15315487.

—-1
—0
—+1

Tyler Vigen. Spurious correlations: Correlation does not imply causation, 2015. URL https://www.tylervigen.com/spurious-correlations.

M. Vojnovic. Contest theory: Incentive mechanisms and ranking methods. 2016.

John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. 2007. doi: 10.2307/3610940.

Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip H.S. Torr. Fast online object tracking and segmentation: A unifying approach. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2019-June, 2019a. doi: 10.1109/CVPR.2019.00142.

Yi Wang, Ning Zhang, Yushi Tan, Tao Hong, Daniel S. Kirschen, and Chongqing Kang. Combining Probabilistic Load Forecasts. *IEEE Transactions on Smart Grid*, 10(4), 2019b. ISSN 19493053. doi: 10.1109/TSG.2018.2833869.

Ziyu Wang, Tom Schaul, and Matteo Hessel. Dueling Network Architectures for Deep Reinforcement Learning. *Advances in Neural Information Processing Systems*, 2017-Decem(2016), 2016. ISSN 10495258.