Forecasting of Political Conflict: Right Data, not Big Data

Philip A. Schrodt, Ph.D.

Parus Analytics LLC and Open Event Data Alliance
Charlottesville, Virginia USA
`http://philipschrodt.org`
`https://github.com/openeventdata/`

Presented at "Can big data enable early action?"
Permanent Mission of Australia to the United Nations, New York
14 March 2018

Well, this is timely...

ESSAYS

Predicting armed conflict: Time to adjust our expectations?

Lars-Erik Cederman[1,*], **Nils B. Weidmann**[2,*]

+ Author Affiliations

*Corresponding author. Email: lcederman@ethz.ch (L.-E.C.); nils.weidmann@uni-konstanz.de (N.B.W.)

Science 03 Feb 2017:
Vol. 355, Issue 6324, pp. 474-476
DOI: 10.1126/science.aal4483

ESSAYS

Bringing probability judgments into policy debates via forecasting tournaments

Philip E. Tetlock[1,*], **Barbara A. Mellers**[1], **J. Peter Scoblic**[2]

+ Author Affiliations

*Corresponding author. Email: tetlock@wharton.upenn.edu

Science 03 Feb 2017:
Vol. 355, Issue 6324, pp. 481-483
DOI: 10.1126/science.aal3147

Prediction and explanation in social systems

Jake M. Hofman[*], **Amit Sharma**[*], **Duncan J. Watts**[*]

+ Author Affiliations

*Corresponding author. Email: jmh@microsoft.com (J.M.H.); amshar@microsoft.com (A.S.); duncan@microsoft.com (D.J.W.)

Science 03 Feb 2017:
Vol. 355, Issue 6324, pp. 486-488
DOI: 10.1126/science.aal3856

And in the *Washington Post*

Monkey Cage | Analysis

Where are coups most likely to occur in 2017?

By Michael D. Ward and Andreas Beger January 31

Supporters of Tukish President Tayyip Erdogan celebrate after soldiers involved in the failed coup attempt surrendered on the Bosporus Bridge in Istanbul on July 16, 2016. (Yagiz Karahan/Reuters)

And from nearby in the current issue of *CACM*

COMMUNICATIONS OF THE ACM

Search

HOME | CURRENT ISSUE | NEWS | BLOGS | OPINION | RESEARCH | PRACTICE | CAREERS | ARCHIVE | VIDEOS

Home / Magazine Archive / March 2018 (Vol. 61, No. 3) / Computational Social Science ≠ Computer Science +... / Full Text

VIEWPOINT

Computational Social Science ≠ Computer Science + Social Data

By Hanna Wallach
Communications of the ACM, Vol. 61 No. 3, Pages 42-44
10.1145/3132698
Comments

VIEW AS: SHARE:

Credit: Evannovostro

This viewpoint is about differences between computer science and social science, and their implications for *computational* social science. Spoiler alert: The punchline is simple. Despite all the hype, machine learning is not a be-all and end-all solution. We still need social scientists if we are going to use machine learning to study social phenomena in a responsible and ethical manner.

I am a machine learning researcher by training. That said, my recent work has been pretty far from traditional machine learning. Instead, my focus has been on computational social science—the study of social phenomena using digitized information and computational and statistical methods.

For example, imagine you want to know how much activity on websites such as Amazon or Netflix is caused by recommendations versus other factors. To answer this question, you might develop a statistical model for

SIGN IN for Full Access

User Name

Password

» Forgot Password?
» Create an ACM Web Account

SIGN IN

ARTICLE CONTENTS:

The Necessity of Prediction in Policy

Feedforward: policy choices must be made in the present for outcomes which may not occur for many years

Planning: even responses to current conditions may require lead times of weeks or months. The typical "policy relevant forecasting interval" is 6 to 24 months.

Factors encouraging technical political forecasting

- Conspicuous failures of existing methods: end of Cold War, post-invasion Iraq, Arab spring
- Success of forecasting models in other behavioral domains
  - Macroeconomic forecasting [maybe...]
  - Elections: Nate Silver (2012) effect
  - Demographic and epidemiological forecasting
  - Famine forecasting: USAID FEWS model
  - Example: statistical models for mortgage repayment in pre-2007 period were quite accurate even if the unpleasant implications were ignored
- Technological imperatives
  - Increased processing capacity
  - Information available on the web
- Decision-makers now expect visual displays of analytical information, which in turn requires systematic measurement
  - "They won't read things any more"

Large Scale Conflict Forecasting Projects

- State Failures Project 1994-2001
- Joint Warfare Analysis Center 1997
- FEWER [Davies and Gurr 1998]
- Center for Army Analysis 2002-2005
- Swiss Peace Foundation FAST 2000-2008
- Political Instability Task Force (PITF) 2002-present
- DARPA Integrated Conflict Early Warning System (ICEWS) 2007-present
- IARPA ACE and OSI
- Peace Research Center Oslo (PRIO) and Uppsala University UCDP models
- EU Joint Research Center Global Conflict Risk Index

Is political behavior predictable? Yes! Good Judgment Project (Tetlock, Meller et al)

- Evaluated about 2000 forecasts, typically with a 6 to 12 month window, across a wide variety political and economic domains
- Most forecasters—more than 90%—were simply "dart-throwing chimps"
- "Superforecasters", however, consistently were about 80% to 85% accurate. This held across multiple years: unlike managed mutual funds, it did not regress to the mean
- Teams of superforecasters were more effective than individuals, and behaved differently than random teams
- Superforecasters have distinct psychological profiles: "foxes rather than hedgehogs"

Superforecaster accuracy is similar to that of a variety of statistical and machine learning models. 80% to 85% appears to be the forecasting "speed limit" in this time frame.

But what about free will?!?

This is relevant to *individual* behavior but is constrained in political behavior

- structural limitations: the Maldives will not respond to climate-induced sea level rise by building a naval fleet to conquer Singapore.
- while individuals *can* change, most of the time they don't
- most politically significant actions involve collective action
- forecasting models predict aggregates, not individual events

While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant.
Sherlock Holmes in *The Sign of the Four*, chapter 10 (1890)

Convergent Results from Forecasting Projects-1

- Most models require only a [very] small number of variables to achieve “speed limit” accuracy
- Indirect indicators—famously, infant mortality rate as an indicator of state capacity in the Political Instability Task Force models—are very useful
- Measurement error on many of the variables being predicted—for example casualties, coup attempts—is still very large.
- Temporal autoregressive effects (repeated behaviors) are huge: the challenge is predicting onsets and cessations, not continuations
- Spatial autoregressive effects—“bad neighborhoods”—are also huge
- Multiple modeling approaches generally converge to similar accuracy

Convergent Results from Forecasting Projects-2

- 80% to 85% accuracy in the 6 to 24 month forecasting window occurs with remarkable consistency: few if any replicable models exceed this, and models below that level can usually be improved. Any claims of greater accuracy out-of-sample should be regarded with extreme skepticism.

- Well-understood open source methods are quite sufficient for these problems: Any proprietary models should be regarded with even more extreme skepticism.

- In all models, there is a tradeoff between the "sensitivity" of a model—how likely is it to give false alarms—and the "recall"—how likely is it to get all of the cases.

- Forecast accuracy does not decline very rapidly with increased forecast windows, suggesting long term structural factors rather than short-term "triggers" are dominant. Trigger models more generally do poorly except as *post hoc* "explanations."

Why have predictive models improved?

Data!


The Economist
Theresa May v Brussels
Ten years on: banking after the crisis
South Korea's unfinished revolution
Biology, but without the cells
The world's most valuable resource
Google
Data and the new rules of competition

PRIO

ACLED Armed Conflict Location & Event Dataset

Minorities at Risk

CORRELATES OF WAR

UPPSALA UNIVERSITET

PENNSTATE

1855

UCDP GEOREFERENCED EVENT DATASET

CIRI Human Rights Data Project

www.humanrightsdata.org

The Dataverse Network™ Project

An Open-Source Application for Publishing, Citing and Discovering Research Data

Polity IV

GTD

Global Terrorism Database

Universität Konstanz

Computing Power

Control Data Corporation 3600
(ca.1965)
32 K (48-bit) RAM memory
1 processor
~1-million operations per second
Output: line printer

Penn State High Performance Computing Facility
15 cluster computers
100 to 2000 2.66 Ghz processors in each cluster
~50 Gb RAM accessible to each processor
130 Tb disk space
4 interactive visualization rooms

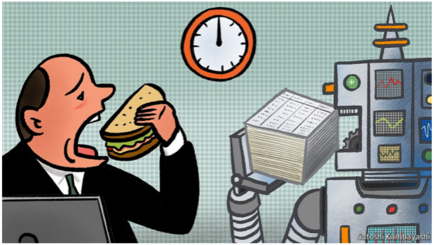
Motorola Razr
16 Gb RAM memory
Dual processor
~500-million operations per sec
540 x 860 color display

New computationally-intensive methods

Unshackled algorithms

Machine-learning promises to shake up large swathes of finance

In fields from trading to credit assessment to fraud prevention, machine-learning is advancing



Print edition | Finance and economics ›
May 25th 2017

MACHINE-LEARNING is beginning to shake up finance. A subset of artificial intelligence (AI) that excels at finding patterns and making predictions, it used to be the preserve of technology firms. The financial industry has jumped on the bandwagon. To cite just a few examples, “heads of machine-learning” can be found at PwC, a consultancy and auditing firm, at JP Morgan Chase, a large bank, and at Man GLG, a hedge-

Open Source Software

MySQL
python
mozilla Firefox
ubuntu
sourceforge
TR
perl
Ruby PROGRAMMING Language
R
WordPress.com
APACHE HTTP SERVER
LibreOffice The Document Foundation
GNU run free run GNU
stackoverflow
FSF FREE SOFTWARE FOUNDATION

Open Event Data Alliance software

Birdcage
Basic, Integrated, and Reliably Distributed Coding, Actors, and Geolocation for Events

PETRARCH family of automated event data coders and dictionaries for CAMEO ontology

PLOVER Event Data Ontology

mordecai
full text geoparsing
github.com/openeventdata/mordecai

FJOLTYNG: PLOVER- and universal dependency-based event coder

ELDIABLO

PETRARCH-based web scraping and event coding pipeline

Is this a big data problem?

Classic definition of “big data”: variety, volume, velocity

- Variety: this we have
- Volume: not so much compared to on-line retailers or medical systems. Most of political conflict events are very rare and occur in a very small number of very idiosyncratic places.
- Velocity: policy-relevant models rarely need true real time data, and often use structural data at the nation-year level. Models can be refined and studied; they do not need to operate in milliseconds.

In addition, we have theories, not just data mining: Amazon [probably] does not have a ”theory of backpacks” even if it sells a lot of them. Substantive understanding remains important.

The Amazon/Google/Alibaba Theory of Backpacks

Brought to you by Big Data and the millions of people who purchase backpacks on-line every year

- If it is August and we have ascertained you are a parent with school-age children, show advertisements for small backpacks
- If it is May and we have ascertained you are between the ages of 18 and 25, show advertisements for large backpacks
- Otherwise show some other advertisement
  - When I was originally preparing this slide in Berlin using Google Docs, I began seeing ads for SAS's machine-learning software. Seriously. Big Data is Watching You!

Apply this approach to conflict, and I'm guessing Thucydides, Machiavelli, and T.R.Gurr still don't have much to worry about

Do we have *too much* data/variety

World Development Indicators projects has 1500+ indicators available!

Advantages of variety (Kraay, WDI)

- Composites have greater stability
- Variance in the measurement provides useful information
- Less affected by biases or methodological weaknesses in individual providers
- Multiple independent sources probably give greater confidence

The world is better connected with information than materially: current Afrobarometer shows only 50% of that population has access to reliable electricity or paved roads, but 92% have access to a cell phone.

Do we have *too much* data/variety?

Disadvantages of variety

- Cost and effort
- Some methods for creating composites aren't transparent or unique
- Weak sources introduce noise and may be systematically biased (e.g. bots in social media)
- When secondary sources are used to generate the original indicator, those aren't actually independent

Historically, the most robust social science models have used only a small number of easily-measured variables, which is quite a different approach than current "big data" models but has a very long and distinguished record (Kahneman)

Simple models are good!

Recent study on predicting criminal recidivism showed equivalent results could be obtained from:

- A proprietary 137-variable black-box system costing $22,000 a year
- Humans recruited from Amazon's Mechanical Turk and provided with 7 variables
- A two-variable statistical regression model

For this problem, there is a widely-recognized "speed limit" on predictive accuracy of around 70% and, as with conflict forecasting, multiple methods can achieve this.

Source: *Science* 359:6373 19 Jan 2018, pg. 263; the original research is reported in *Science Advances* 10.1126/sciadv.aao5580 (2018)

Political Instability Task Force operational modeling approach

- Accumulate a large number of variables from open sources and exhaustively explore combinations of these using a variety of statistical and machine-learning approaches: this establishes the out-of-sample "speed limit"
- The "speed limit" should be similar to the accuracy of human "super-forecasters" (Tetlock)
- Construct operational models with "speed limit" performance using very simple sets of variables—typically about five—using the most robustly measured of the relevant independent variables

Simple models are transparent; robust measures are transparent and inexpensive

Some challenges

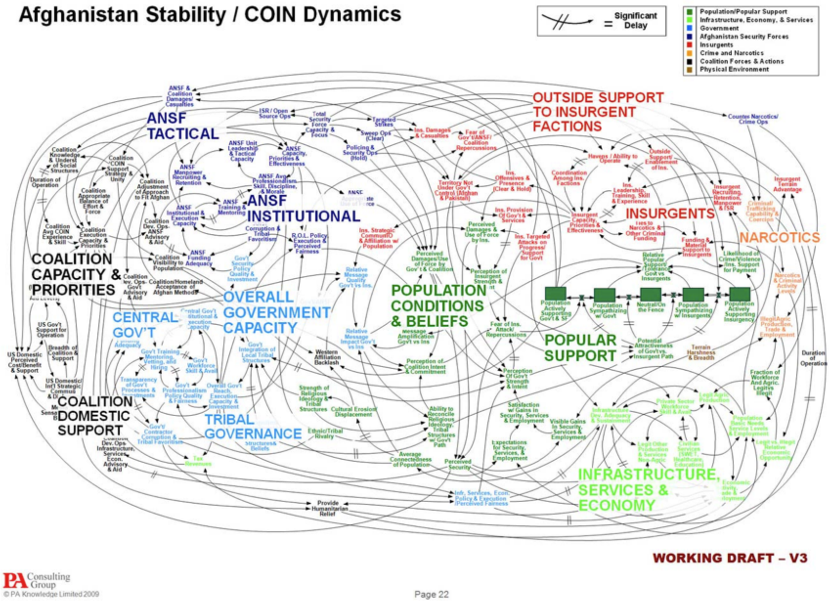
Challenges applying this to foreign policy

- Integrating quantitative analysis into traditionally qualitative decision-making
- Economic historians have found that efficiently integrating a new technology (e.g. steam power; electricity; computers) into an industry takes about 20 years, a human generation
- Rare events and probability analysis are difficult for everyone, including statisticians (Kahneman)
  - Questions such as the relationship between climate change and conflict are *very* difficult to study and we won't have immediate answers
- Visualization is also difficult (Tufte): machine-assisted self-deception
- Political sensitivity: transparency might help here

Irreducible sources of error-1

- Specification error: no model of a complex, open system can contain all of the relevant variables;

Open, complex systems



Irreducible sources of error-1

- Specification error: no model of a complex, open system can contain all of the relevant variables;
- Measurement error: with very few exceptions, variables will contain some measurement error
  - presupposing there is even agreement on what the "correct" measurement is in an ideal setting;
  - This biases the estimation of the model as well as the predictions
- Quasi-random structural error: Complex and chaotic deterministic systems behave as if they were random under at least some parameter combinations. Chaotic behavior can occur in equations as simple as $x_{t+1} = a{x_t}^2 + bx_t$

Irreducible sources of error-2

- Rational randomness such as that predicted by mixed strategies in zero-sum games
- Arational randomness attributable to free-will
  - Rule-of-thumb from our rat-running colleagues: "A genetically standardized experimental animal, subjected to carefully controlled stimuli in a laboratory setting, will do whatever it wants."
- Effective policy response: in at least some instances organizations will have taken steps to head off a crisis that would have otherwise occurred.
- The effects of natural phenomenon
  - the 2004 Indian Ocean tsunami dramatically reduced violence in the long-running conflict in Aceh

(Tetlock (2013) independently has an almost identical list of the irreducible sources of error.)

Balancing factors which make behavior predictable

- Individual preferences and expectations, which tend to change very slowly
- Organizational and bureaucratic rules and norms
- Constraints of mass mobilization strategies
- Structural constraints
- Choices and strategies at Nash equilibrium points where no actor has an incentive to change
- Organizations and individuals have a strong tendency to just repeat what they've already been doing
- Network and contagion effects (same)

"History doesn't repeat itself but it rhymes"
(variously attributed to Mark Twain and Friedrich Nietzsche; neither said it)

Theory matters (see Wallach, *CACM* March-2018)

Arab Spring is an unprecedented product of the new social media

- Model used by Chinese censors of new social media: King, Peng, Roberts 2012
- Next likely candidates: Africa

Arab Spring is an example of an instability contagion/diffusion process in a system already structurally primed for change

- Eastern Europe 1989-1991, OECD 1968, US South 1859-1861, Europe 1848, Latin America 1820-1828
- Next likely candidates: Central Asia

Arab Spring is a black swan

- There is no point in modeling black swans, you instead build systems robust against them

Paradox of political prediction

Political behaviors are generally highly incremental and vary little from day to day, or even century to century (Putnam).

Nonetheless, we *perceive* politics as very unpredictable because we are adapted by evolution to focus on the unexpected (Kahneman, Pinker).

Consequently the only "interesting" forecasts are those which are least characteristic of the system as a whole. However, only some of those changes are actually predictable.

Ethical concerns

- Thus far, we've generally had the luxury of no one paying attention to any of our predictions : what if governments do start paying attention?
  - USAID/FAO famine forecasting model is a positive example of this
  - It is *possible* that our models could become less accurate because crises are being averted, but I don't see that happening any time soon.
- Difficulties in getting *anyone*, including experts (see Kahneman, Tetlock), to correctly interpret probabilistic forecasts
- Possible impact on sources
  - Local collaborators
  - Journalists (cf. Mexico)
  - NGOs to the extent we are using their information

Thank you

Email:
schrodt735@gmail.com

Slides:
`http://eventdata.parusanalytics.com/presentations.html`

Links to data and software:
`http://openeventdata.org/`
`https://github.com/openeventdata/`

Blog: `http://asecondmouse.org`

Supplementary Slides

Challenge: distinguishing black swans from rare events

Black swan: an event that has a low probability even conditional on other variables

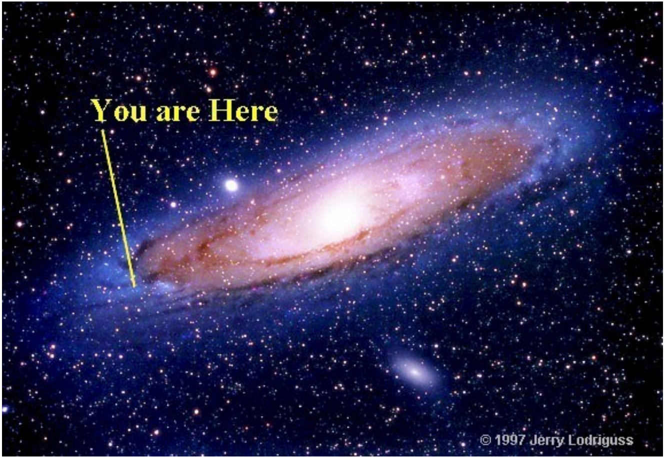
Rare event: an event that occurs infrequently, but conditional on an appropriate set of variables, does not have a low probability

Medical analogy: certain rare forms of cancer appear to be highly correlated with specific rare genetic mutations. Conditioned on those mutations, they are not black swans.

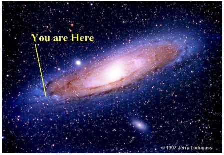
Another important category: high probability events which are ignored. The "sub-prime mortgage crisis" was the result of the failure of a large number of mortgages which models had completely accurately identified as "sub-prime" and thus likely to fail. This was not a low probability event.
Upton Sinclair: It is hard to persuade someone to believe something when he can make a great deal of money not believing it.

Black swans

Ideal forecasting targets are neither too common nor too rare



Finding a non-trivial forecast



- Too frequent: prediction is obvious without technical assistance
- Too rare: prediction may be correct, but the event is so infrequent that
  - The prediction is irrelevant to policy
  - Calibration can be very tricky
  - Accuracy of the model is difficult to assess
- "Just right": these are situations where typical human accuracy is likely to be flawed, and consequently these could have a high payoff.

Differing attitudes towards error

Geography, physics:

- Progress consists of ever more accurate data

Political science:

- Trust nothing—everything has error, just control for the systematic biases

Machine learning::

- it is what it is: goal is improving prediction

Statistics::

- signal to noise: Perfect is the enemy of "good enough"
- mathematically approximate the characteristics of the error
- Taleb, Mandelbrot: don't be a Gaussian in a power-law world

Statistical and modeling challenges

Rare events

- Incorporate much longer historical time lines?—Schelling used Caesar's *Gallic Wars* to analyze nuclear deterrence
- New approaches made possible by computational advances

Analysis of event sequences, which are not a standard data type

- There are, however, a large number of available methods, and it is just possible that these will work with very large data sets

Causality

- Oxford *Handbook of Causation* is 800 pages long

Integration of qualitative and qualitative/subject-matter-expert (SME) information

- Bayesian approaches using prior probabilities are promising but to date they have not really been used

Computationally-intensive methods

- "deep learning" neural networks, the current hot item in machine learning
- Bayesian estimation using Markov chain Monte Carlo methods
- Bayesian model averaging ("*AJPS*-as-algorithm")
- random forest models
- large-scale textual databases
- machine translation
- geospatial visualization
- real-time automated coding
- remote sensing data such as nightlight density

The very finite set of widely used ML methods

- Support vector machines
- Clustering, typically using k-means
- Random forests, a relatively recent ensemble variation on the older method of decision trees
- Neural networks
  - A very old method which is now being used with vastly greater hardware and a few new algorithmic tricks to create “deep learning”
- Genetic algorithms
- Logistic regression, which not infrequently is ”embarrassingly effective”

New opportunities from machine learning

- ML methods recently have been successful in a number of "artificial intelligence" problems previously thought to be unsolvable
- Most statistical models have already been extensively explored, and in any case are not optimized for prediction (Ward, Greenhill and Bakke 2010)
- The parameter spaces of many of these models are vastly larger than those of statistical models
- ML models generally work well with heterogeneous cases
- Most ML models are relatively insensitive to missing values, or treat it as information
- Software is readily available and open source

Risks in machine learning models

- Over-fitting
- It is not clear that political conflict early warning has a sufficient number of cases to take advantage of methods which require large amounts of data
- ML models are generally atheoretical, and the rich parameter spaces mean it is often difficult to impossible to ascertain the relative importance of independent variables
- Some models—notably "deep learning"—are quite new and may have features we don't fully understand
- In many instances, ML models show only marginal improvements over well-understood methods such as logistic regression when applied across a wide set of out-of-sample problems

Some interesting open questions

- Under what circumstances does climate change increase versus reduce conflict?
  - Contrary to the ubiquitous “Battle at the water hole” analogies, there is ample evidence to support both effects
- How can event data and structural data be combined to increase predictive accuracy?: to date, they largely just seem to be substitutable
- Are “trigger models” real or simply a cognitive illusion?
- How many theoretically distinct forms of sub-state conflict should be analyzed?
- What is the optimal level of detail in event data and geospatial data (which will depend on the question, of course)