Technical Forecasting of Political Conflict

Philip A. Schrodt

Parus Analytics LLC
Charlottesville, VA
http://philipschrodt.org
http://eventdata.parusanalytics.com

DIA Community of Practice
19 January 2021

PARUS
ANALYTICS

Well, someone is paying attention...

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)

Foreign Policy

▶ ARGUMENT PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

Why the World Can't Have a Nate Silver

The quants are riding high after Team Data crushed Team Gut in the U.S. election forecasts. But predicting the Electoral College vote is child's play next to some of these hard targets.

BY JAY ULFELDER | NOVEMBER 8, 2012

Vs.

▶ ARGUMENT PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

Predicting the Future Is Easier Than It Looks

Nate Silver was just the beginning. Some of the same statistical techniques used by America's forecaster-in-chief are about to revolutionize world politics.

BY MICHAEL D. WARD , NILS METTERNICH | NOVEMBER 16, 2012

Two approaches that did not work well in the past

Qualitative

Quantitative

Problems with qualitative approaches

Tetlock: Experts typically do about as well as a "dart-throwing chimp"

Problems with qualitative approaches

Tetlock: Experts typically do about as well as a "dart-throwing chimp"

Except for television pundits, who do even worse. As did pretty much every political expert involved in forecasting the 2016 and 2020 US presidential elections.

The media want things to be dramatic. "We're all going to die! Details at 10!!!"

Problems with qualitative approaches

Tetlock: Experts typically do about as well as a "dart-throwing chimp"

Except for television pundits, who do even worse. As did pretty much every political expert involved in forecasting the 2016 and 2020 US presidential elections.

The media want things to be dramatic. "We're all going to die! Details at 10!!!"

Qualitative *theory* isn't much better:
Remember the hegemonic US seizure of undefended Canadian and Mexican oil fields in response to the 1973 OPEC oil embargo?

Neither do I.

SMEs and the "narrative fallacy"

SME = "subject matter expert"

Hegel: the owl of Minerva flies only at dusk

Taleb (*Black Swan*): seeking out narratives is an almost unavoidable cognitive function and it generates a dopamine hit

Tetlock (Good Judgement Project): prior knowledge as a SME contributes only 2% to improved forecasting accuracy

This is your brain on narratives

Problems with quantitative approaches

Ward, Greenhill and Bakke (2010): Models based on significance tests don't predict well because that is not what a significance test is supposed to do.

Gill, Jeff. 1999. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly* 52:3, 647-674.

Problems with quantitative approaches

Ward, Greenhill and Bakke (2010): Models based on significance tests don't predict well because that is not what a significance test is supposed to do.

Gill, Jeff. 1999. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly* 52:3, 647-674.

Frequentism is logically inconsistent and has been characterized in Meehl (1978) as "a terrible mistake, basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology"

- Hey, dude, tell us what you *really* think...
- But that is another lecture...

Kahneman et al: people are really bad at statistical reasoning

- Everyone, including statisticians unless they focus very hard
- Example: managed mutual funds, which both theory and evidence indicate cannot work (and thus are rapidly disappearing)
- Example: opposition to "evidence based medicine" in the US, with a preference for clinical intuition even when this has been demonstrated to be less effective
- Probabilitistic weather forecasts seem to be the one major exception: rain likelihood, hurricane tracks

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—about 90%—were simply "dart-throwing chimps"
- "Superforecasters", however, consistently were about 80% to 85% accurate, similar to that of the PITF and ICEWS models. 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"
- Prediction markets, SMEs and ensemble models provided only marginal improvements

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 Times: even responses to current conditions may require lead times of weeks or months

Wayne Gretzky Principle: You skate to where the puck is going to be, not where it is.

And it might go a lot deeper than that...

Lisa Feldman Barrett. *Seven and a Half Lessons About the Brain* 2020:

Pattern-based prediction is the *fundamental* purpose of the cognitive system: the evolutionary advantages are simply too great, both in terms of avoiding predators and being a predator, and in allocating a finite energy budget. This got going during the Cambrian period.

In most mammals, and certainly primates, this extends to evaluating and positioning oneself in hierarchies and accounts for a great deal of cognitive effort. (Robert M. Sapolsky. *Behave: The Biology of Humans at Our Best and Worst*, 2017)

Hey, hold on a moment. . .
what do you mean by "forecast"

(or "prediction" for that matter)

Classical forecasting

Predict a quantity at a known point in the future (preferably well in advance)

- Stocks and other economic indicators
- Election results
- Sports outcomes (really big business!)

Predict whether something will occur within some time frame

- Almost all political conflict forecasting problems
- Credit card and tax return fraud (will the transaction later be found to be problematic?)
- Advertising effectiveness (based on historical data, will an ad placement increase sales?)

In both, the forecast is often in terms of probabilities

Contemporary new forecasting approaches

Unbelievably huge business, though mostly not relevant to political forecasting

- "Algorithmic" financial trading at micro-second intervals, in contrast to long time frames
- Selecting content which is most likely to keep a user on a site (Netflix; YouTube)
- Optimally selecting on advertisements on micro-second time frames

"Forecast" versus "Prediction"

This is just the good old parallel Germanic ("forecast") and Latin ("prediction") roots found all over the English language:

- breaking and entering
- cease and desist
- last will and testament

Some people wax fanatically that these are distinct but, as with much of this field (cf. totally different terms applied to the same concepts in statistics versus machine learning: "logistic regression" vs. "entropy maximization") distinct meanings have not been standardized.

Meh. . . whatever. . . sorry. . .

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 prior to about 2015: Nate Silver effect
  - Demographic and epidemiological forecasting
  - Famine forecasting: USAID FEWS model
  - Example: statistical models for mortgage repayment were quite accurate
- Technological imperatives
  - Increased processing capacity
  - Increased variety of predictive models, all open source
  - 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"

This must be important: it's in *The Economist*!

The science of civil war

What makes heroic strife

Computer models that can predict the outbreak and spread of civil conflict are being developed

Apr 21st 2012 | from the print edition

And *The Chronicle*!

THE CHRONICLE OF HIGHER EDUCATION
November 5, 2014

The Chronicle Review

October 13, 2014

Can We Predict the Next War?

Mark Shaver for The Chronicle Review

Enlarge Image

By Beth McMurtrie

Every minute of every day, computers overseen by Virginia Tech process billions of bits of data in an attempt to predict the future.

Tweets from politicians, satellite images of hospital parking lots, news stories about rising bus fares: All are mined, categorized, and fed into algorithms designed to anticipate the next flu outbreak or which candidate will win a city election. As many as 50 computer-generated alerts flash daily on computer monitors and are evaluated for accuracy at the end of each month.

Most Popular
Most Viewed | Most Comme
1. Seeking a Climate Change
2. What the Midterm Election
3. Outcome of Governors' Rac
Policy in Several States
4. Confessions of a Young, Pr

The Forecaster's Quartet

- Nassem Nicholas Taleb. *The Black Swan*
  (most ~~entertaining~~ obnoxious)
- Daniel Kahneman. *Thinking Fast and Slow*
  (30 years of research which won Nobel Prize)
- Philip Tetlock. *Expert Political Judgment*
  Superforecasting: The Art and Science of Prediction
  (Tetlock and Dan Gardner) (most directly relevant)
- Nate Silver. *The Signal and the Noise*
  (high level of credibility after perfect 2012 electoral vote predictions but most examples are not election forecasts. E.g. major shout-out to *Moneyball*)

Past 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 2002-present
- DARPA ICEWS 2007-2015
- IARPA ACE and OSI ca 2015
- Peace Research Center Oslo (PRIO) and Uppsala University UCDP models
- US Holocaust Memorial Museum Prediction Poll

Forecasting Projects

Other policy-oriented groups who have presented models in various venues over the past five years but generally don't publicize these

- Germany
- Netherlands
- World Bank
- ECOWAS

Presumably there are others, some classified. At the nation-year level at least, this is seen as an existing technology that can be customized.

Political Instability Task Force

- US government, multi-agency: 1995-present
- Statistical modeling of various forms of state-level instability
- Forecasting models actively used since about 2005
  - Two year probability forecasts with roughly 80% accuracy (AUC)
  - Predominantly logistic models with a simple "standard PITF"set of variables; shifting to Bayesian approaches
  - (PITF has accumulated a set of 2700 variables but only a small number end up being important predictors)

PITF Variables

Variables Tested

CONCEPT	SELECTED EXAMPLES OF MEASURES TESTED
state capacity	infant mortality, population, GDP, military personnel, polity durability
violent conflict	civil war, armed attacks, regional conflicts, reported fatalities in political violence, government mass killing
non-violent challenges to state authority	protests, strikes, government crises
government institutions	democracy, autocracy, factionalism, other polity measures
ethnic relations	ethnic diversity, elite ethnicity, state-led discrimination
demographics	youth-bulge
international ties	GATT/WTO membership, trade-openness

Two-year time horizon tends to favor structural variables Source: Ben Valentino and Chad Hazlett, "Forecasting Non-state Mass Killings", October 2012

PITF Results, ca. 2005

A Global Model for Forecasting Political Instability

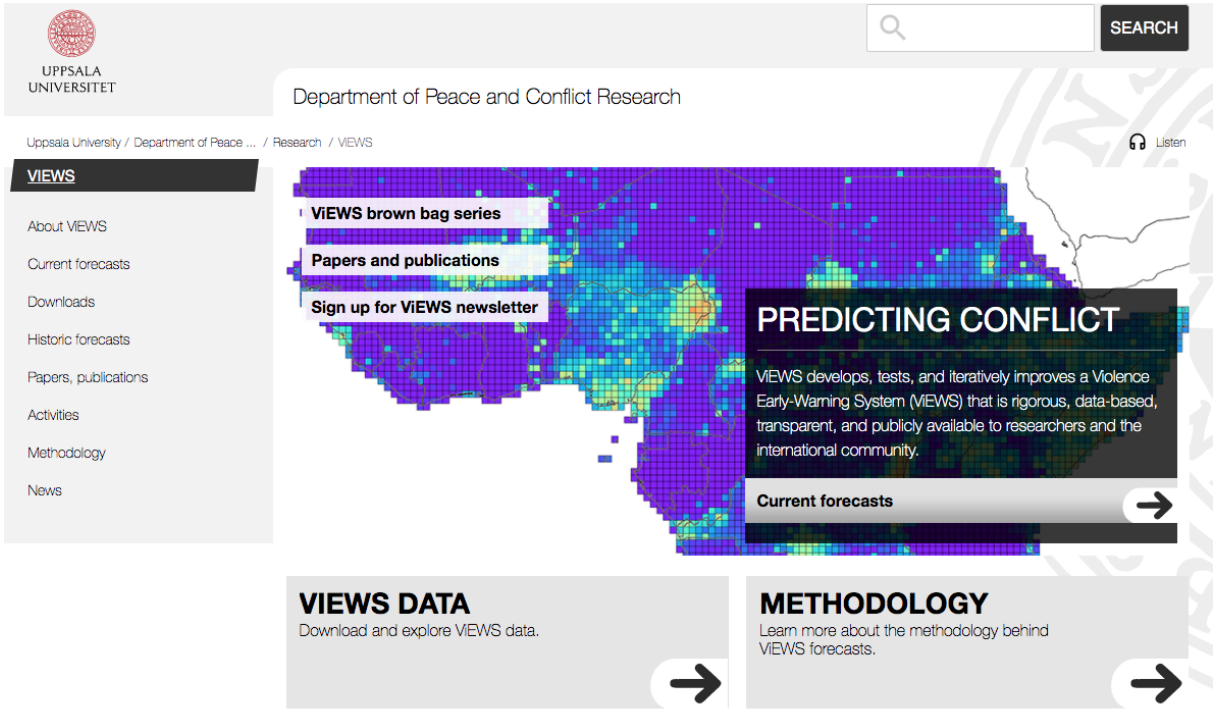
Jack A. Goldstone George Mason University
Robert H. Bates Harvard University
David L. Epstein Columbia University
Ted Robert Gurr University of Maryland
Michael B. Lustik Science Applications International Corporation (SAIC)
Monty G. Marshall George Mason University
Jay Ulfelder Science Applications International Corporation (SAIC)
Mark Woodward Arizona State University

Examining onsets of political instability in countries worldwide from 1955 to 2003, we develop a model that distinguishes countries that experienced instability from those that remained stable with a two-year lead time and over 80% accuracy. Intriguingly, the model uses few variables and a simple specification. The model is accurate in forecasting the onsets of both violent civil wars and nonviolent democratic reversals, suggesting common factors in both types of change. Whereas regime type is typically measured using linear or binary indicators of democracy/autocracy derived from the 21-point Polity scale, the model uses a nonlinear five-category measure of regime type based on the Polity components. This new measure of regime type emerges as the most powerful predictor of instability onsets, leading us to conclude that political institutions, properly specified, and not economic conditions, demography, or geography, are the most important predictors of the onset of political instability.

Source: Amer J of Pol Sci Vol 54, no. 1, Jan 2010 pg. 190

Dynamic Global Conflict Risk Index: European Commission

ViEWS: Uppsala University and Peace Research Institute, Oslo



ViEWS: a political Violence Early-Warning System

77

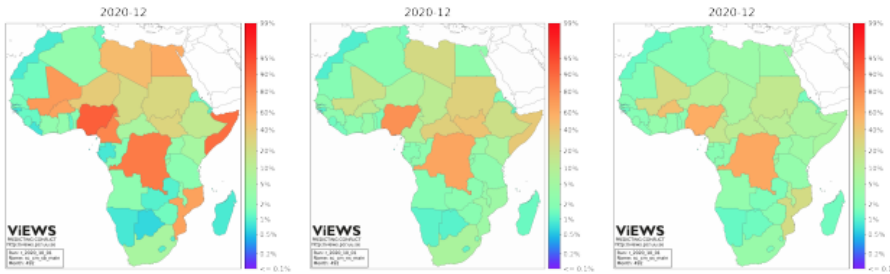
The ViEWS project started in January 2017 with funding from the ERC Advanced Grant scheme. It will develop, test, and iteratively improve a pilot Violence Early-Warning System (ViEWS) that is rigorous, data-based, and publicly available to researchers and the international community. ViEWS will provide early warnings for the three forms of political violence recorded by the UCDP: armed conflict involving states and rebel groups, armed conflict between non-state actors, and violence against civilians, and apply these to specific actors, sub-national geographical units, and countries. Early warnings for forced population displacement will be added as a fourth category. Core tasks of the ViEWS project will be to add detailed geographical information, link to the other levels of analysis in the system, and iteratively improve the forecasts by means of out-of-sample evaluation and Bayesian Model Averaging.

ViEWS: Monthly forecasts

ViEWS monthly forecasts, October 2020*

Summary of forecasts

Friday 30th October, 2020



(a) State-based conflict (sb) (b) Non-state conflict (ns) (c) One-sided violence (os)

Figure 1: Ensemble forecasts for December 2020.

This report presents ViEWS forecasts at $s = 4$ for December 2020 as of 1 October 2020, which are based on data that are updated up to and including August 2020. The underlying conflict data were produced by the UCDP (`http://ucdp.uu.se`). The ViEWS compilation of these data and data from other sources are available at `https://www.pcr.uu.se/research/views/data/downloads/`.

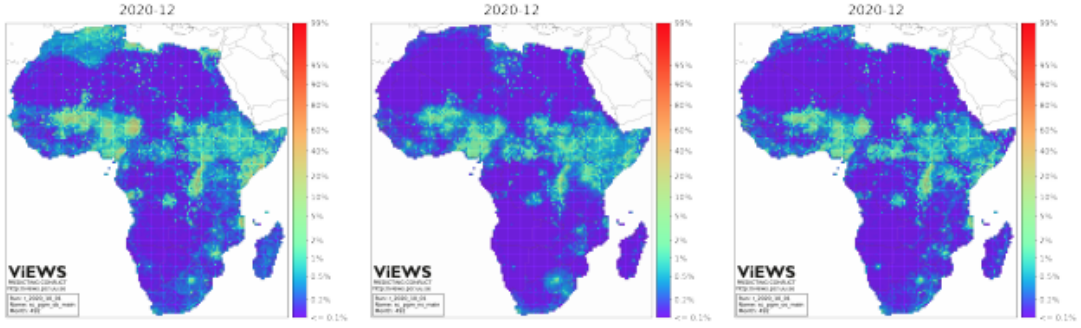
In the following, we highlight developments in the most recent months. For a discussion of what underlies the forecasts in terms of slowly changing risk factors as well as methodological

*This report was prepared by Håvard Hegre, Mihai Croicu, Remco Jansen, Angelica Lindqvist-McGowan, and Maxine Ria Leis. The research was funded by the European Research Council, project H2020-ERC-2015-AdG 694640 (ViEWS). The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

ViEWS: PRIO-GRID resolution for Africa

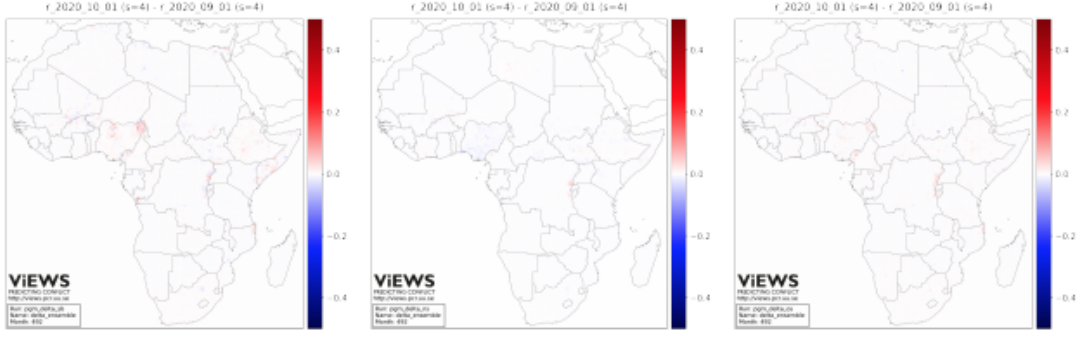
2 PRIO-GRID-month forecasts for December 2020

Figure 5 shows the probability of at least one fatality in December 2020 in each fine-grained sub-national geographical location ('PRIO-GRID cell') and for each of the three outcomes. The forecasts are based on data up to and including August 2020. The color mapping is the same as for the country-month forecasts.



(a) State-based conflict (sb), December 2020 (b) Non-state conflict (ns), December 2020 (c) One-sided violence (os), December 2020

Figure 5: Ensemble forecasts for December 2020



(a) State-based conflict (sb) (b) Non-state conflict (ns) (c) One-sided violence (os)

Figure 6: Change in *pgm* predictions at $s = 4$ compared to last month

Convergent Results from Forecasting Projects-1

- Most models require only a [very] small number of variables
- Indirect indicators—famously, infant mortality rate as an indicator of development—are very useful
- Temporal autoregressive effects 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 sense of AUC around 0.8—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
- Measurement error on many of the dependent variables—for example casualties, coup attempts—is still very large
- 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."

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 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)

Conjecture

For the possibly first time in history, we may be entering an era when foreign policy can be based on relatively accurate projections of the future rather than random guesses and ideology

"Possibly" since the superforecaster approach may have been independently discovered earlier, for example in Confucian and Venetian bureaucracies. Properly done, aggregates of weak forecasts can be used to provide strong forecasts: "boosting" algorithms in machine learning demonstrate this systematically.

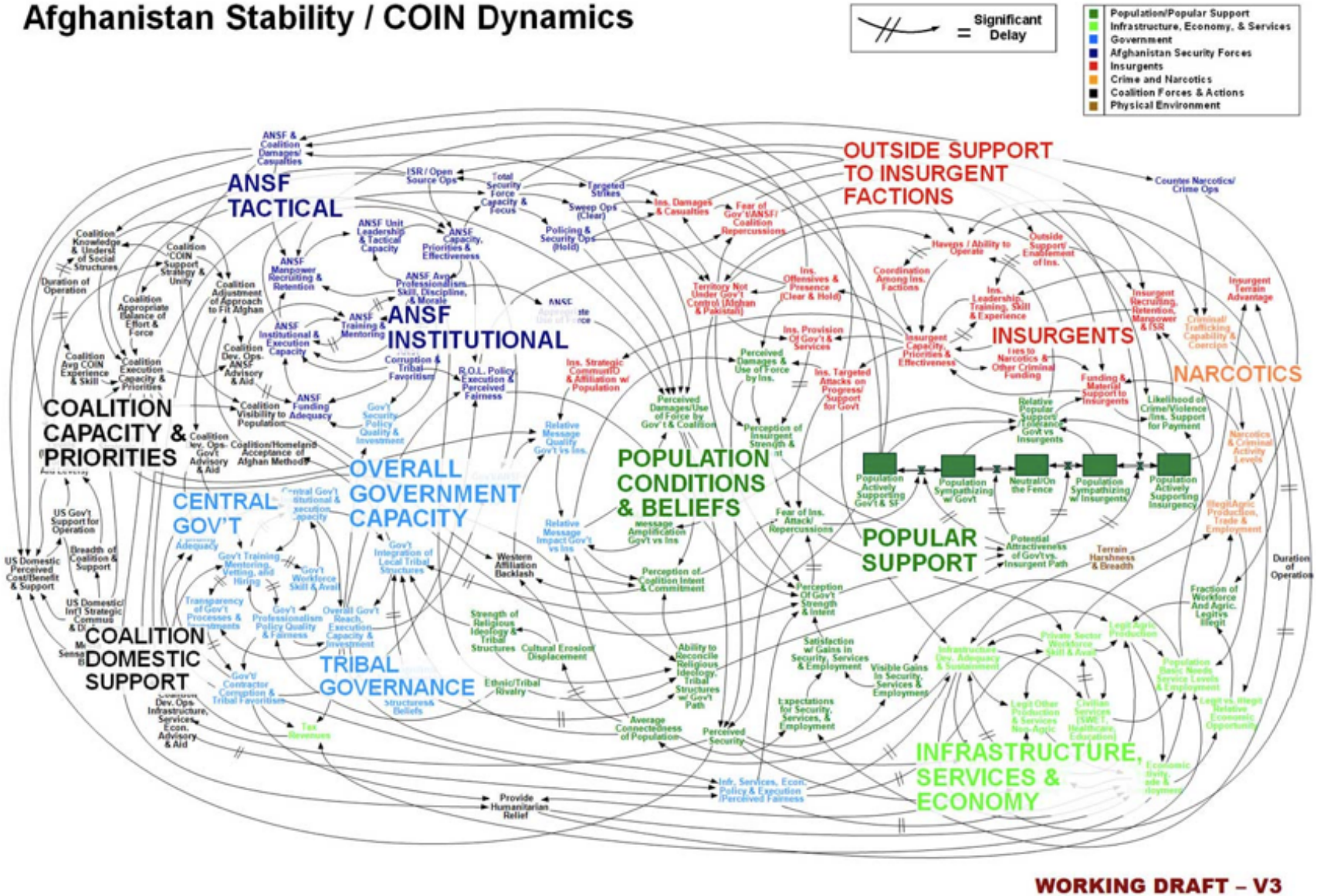
Three other cases where expensive "professional" advice has been random or worse

- Medicine prior to sometime in the 20th century
- Managed mutual funds
- SAT and GRE scores (ouch!)

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;
  - Predictive accuracy is limited by the square root of measurement error: in a bivariate model if your reliability is 80%, your accuracy can't be more than 90%
  - This biases the coefficient estimates 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 $x_{t+1} = a{x_t}^2 + bx_t$

Open, complex systems



Irreducible sources of error-2

- Rational randomness such as that predicted by mixed strategies in zero-sum games
- 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
- 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."

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

Hey, I'm no rat! What about free will?!?

This is relevant to *individual* behavior but is constrained in political—and economic—behavior

- 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)

That said, if Tetlock is correct that 95% of the population are dart-throwing chimps, for them political behavior is unpredictable.

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:
  the Maldives will not respond to climate-induced sea level rise by building a naval fleet to conquer Singapore.
- Choices and strategies at Nash equilibrium points
- Autoregression (more a result than a cause)
- Network and contagion effects (same)

"History doesn't repeat itself but it rhymes"
Mark Twain (also occasionally attributed to Friedrich Nietzsche)

OPPORTUNITIES

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

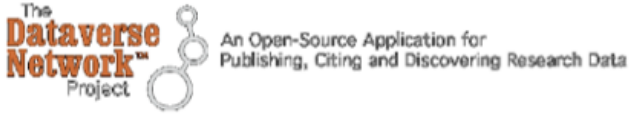

CORRELATES OF
WAR


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

ICEWS event data

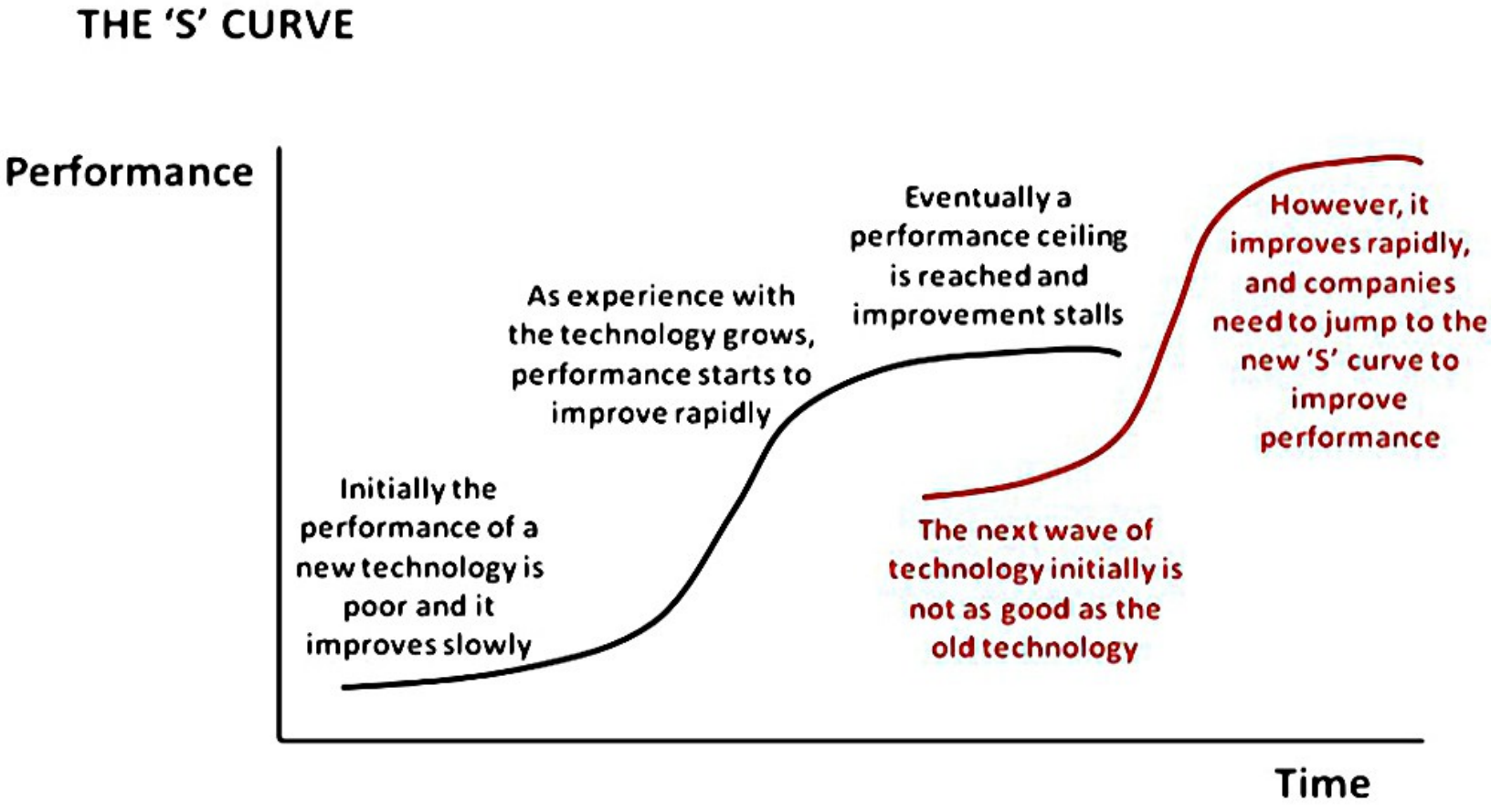
`https://asecondmouse.wordpress.com/2015/03/30/seven-observations-on-the-newly-released-icews-data/`

- Now funded by PITF so probably stable for the near future
- Data released on Dataverse and covers 1996-present with weekly updates
- Government version includes source texts
- Raytheon/BBN [proprietary] Serif/ACCENT coder
- Global coverage but events are still disproportionately from Asia (PITF is working to correct this)
- BBN has extensively refined the CAMEO specification and coding manual is on Dataverse
- Includes a subset of the actor dictionaries but not verb dictionaries
- Geolocated, though with quite a few errors

Weaknesses in the CAMEO coding system

- Based on the Cold War-era WEIS system so state-centered though sub- and non-state actors have been added
- Over 300 categories in a complex hierarchy: this approach was inherited from WEIS
- Developed as a "first draft" for a very specific NSF project on conflict mediation and not intended as a universal system; adopted by the Lockheed ICEWS effort in 2009 as an expedient
- Some major types of political behavior—elections, legislative behavior—are missing. No coding of natural disasters or criminal activity.
- ICEWS implementation is text-oriented and not very statistics/ML friendly
- Most of the new dictionary development was done on a small number of categories; coverage is uneven

Is event data ready for disruption?



Are we at the flat point on a lower S-curve?

- David Honey (DARPA/ODNI) notes that hype is maximized when the curve flattens: please note that at present most people think event data sucks
- Machine coding did a classical disruption on human coding because it was lower quality but cheaper: in Clayton Christensen's theory this drives S-curve disruptions.
- Machine learning classifiers—support vector machines or neural networks—might replace patterns/dictionaries as cheaper-not-better if gold standard records (GSRs) become available. This has been done on toy problems.
- S-curves can level off and stay there:
  - Diesel locomotives
  - Boeing 737
  - 70-mph highway speed limit

Another take on this

- An IARPA PM at recent meeting: "I've talked to lots of analysts: no one has any use for event data."
- Twelve hours later, same meeting, a government analyst: "We *love* your event data tension model!" Suggesting the issue is open
- Observation: Event data never really takes off—in either government or academic research—but it also never goes away: see `http://openeventdata.org/datasets.html` which lists 16 active projects ca. 2018.
- Observation: For the first time in the history of the field, the most innovative work has shifted to Europe—VIEWS, GCRI, ACLED, EMM. These slides are based on talks I've given this year in Berlin and Brussels, not Washington.

PLOVER

Political Language Ontology for Verifiable Event Records

Event, Actor and Data Interchange Specification

Open Event Data Alliance

http://openeventdata.org/

http://ploverdata.org/

DRAFT Version: 0.6b2

March 2017

PLOVER objectives

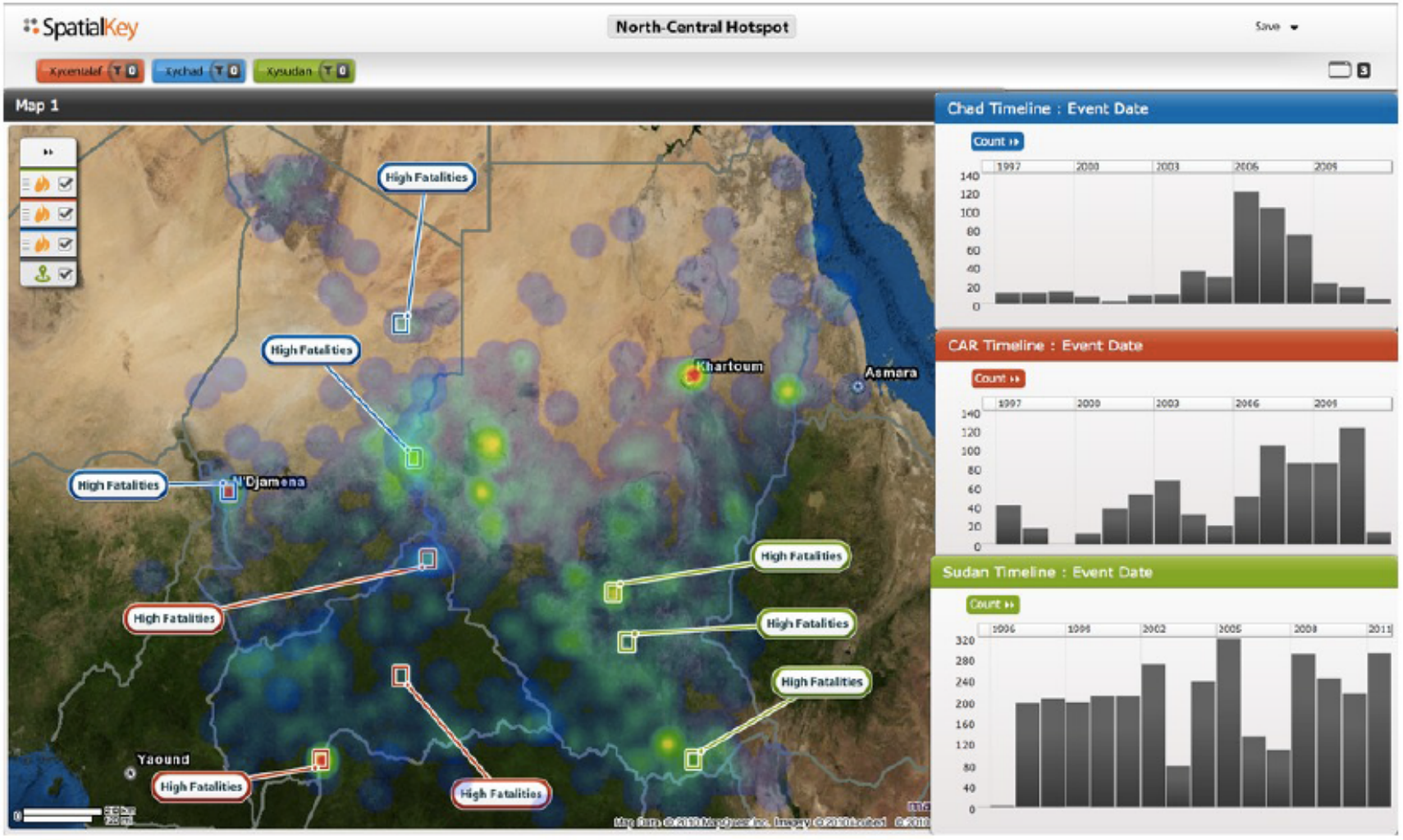
- Only the 2-digit event "cue categories" have been retained from CAMEO. These are defined in greater detail than they were in WEIS and CAMEO.
- Some additional consolidation of CAMEO codes, and a new category for criminal behavior
- Standard optional fields have been defined for some categories, and the "target" is optional in some categories.
- A set of standardized names ("fields") for line-delimited JSON (`http://www.json.org/`) records are specified for both the core event data fields and for extended information such as geolocation and extracted texts;
- We have converted all of the examples in the CAMEO manual to an initial set of English-language "gold standard records" for validation purposes—these files are at `https://github.com/openeventdata/PLOVER/blob/master/PLOVER_GSR_CAMEO.txt`

Event, Mode, and Context

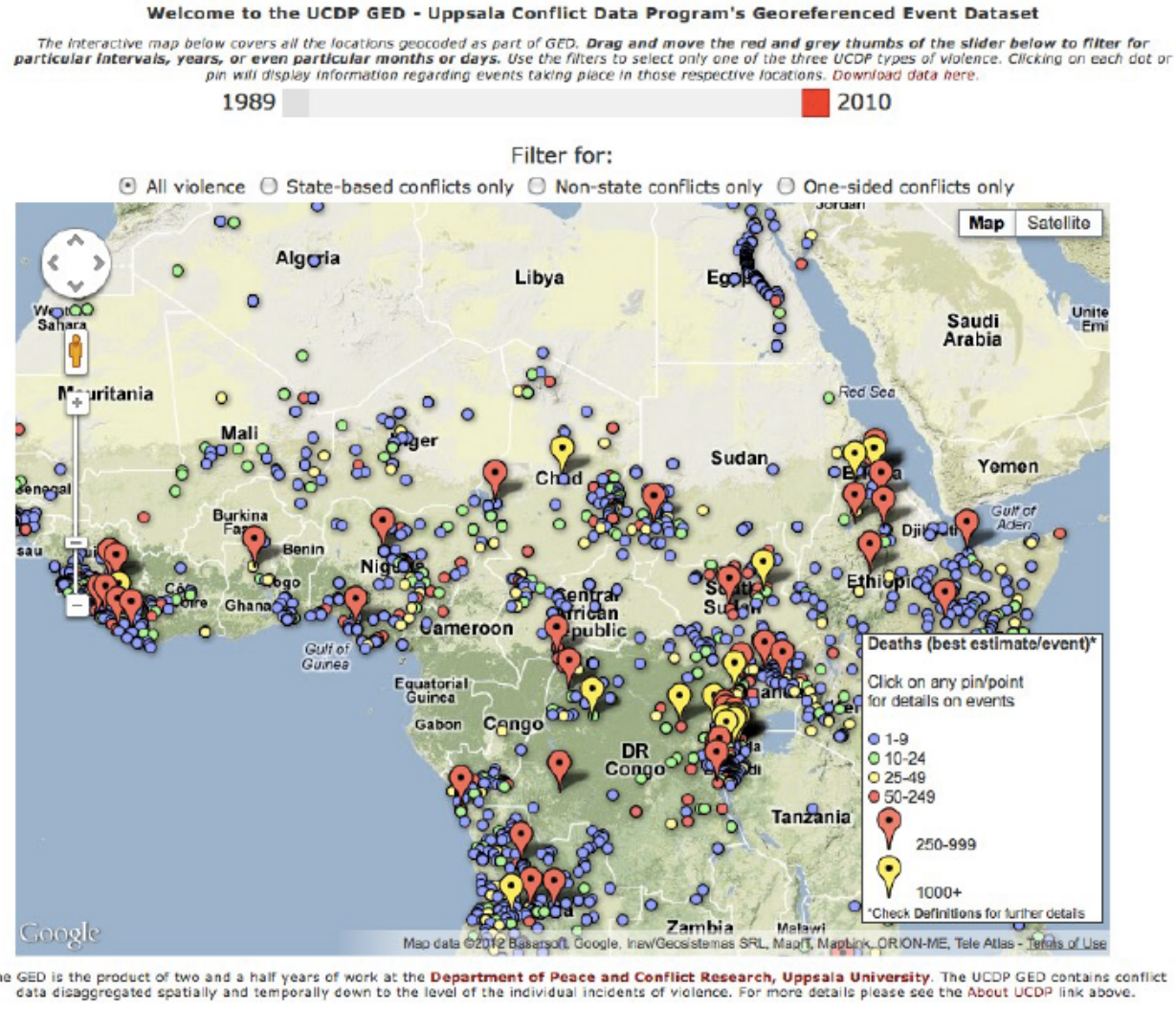
Most of the detail found in the 3- and 4-digit categories of CAMEO is now found in the *mode* and *context* fields in PLOVER. More generally, PLOVER takes the general purpose "events" of CAMEO (as well as the earlier WEIS, IDEA and COPDAB ontologies) and splits these into "$event - mode - context$" which generally corresponds to "$what - how - why$." We anticipate at least four advantages to this:

1. The "$what - how - why$" components are now distinct, whereas various CAMEO subcategories inconsistently used the *how* and *why* to distinguish between subcategories.
2. We are probably increasing the ability of automated classifiers—as distinct from parser/coders—to assign *mode* and *context* compared to their ability to assign subcategories.
3. In initial experiments, it appears this approach is *much* easier for humans to code than the hierarchical structure of CAMEO because a human coder can hold most of the relevant categories in working memory (well, that and a few tables easily displayed on a screen)
4. Because the words used in differentiate *mode* and *context* are generally very basic, translations of the coding protocols into languages other than English is likely to be easier than translating the subcategory descriptions found in CAMEO.

Location: ACLED Geospatial



Location: UCDP Geospatial



The GED is the product of two and a half years of work at the **Department of Peace and Conflict Research, Uppsala University**. The UCDP GED contains conflict data disaggregated spatially and temporally down to the level of the individual incidents of violence. For more details please see the About UCDP link above.

Conjecture: Sub-state, sub-annual models are the future

The problem of predicting at the nation-year level has essentially been solved: it is straightforward to construct models which hit the "speed limit" of 75% to 85% accuracy.

Sub-state and sub-annual predictions are both a new challenge and more relevant to contemporary policy: this is seen explicitly in ViEWS and GCRI is also moving in this direction.

At the nation-year level, there was a good alignment between ends and means: states tend to change policies slowly, and annualized state-level data is very common. But that was then...

Sub-state and sub-annual data possibilities

- I'm obviously biased, but it is hard to imagine that event data produced by fully automated coding—easily resolved to within a day or so, and to a location and/or a specific actor—is not going to be important. However, existing systems use technology that has not been updated with the radical improvements in natural language processing of the past five years.

- Remote sensing data data is now widely available (along with open source software for processing it) and becoming less expensive

- The ACLED and ECOWAS model of local networks [discretely] reporting to IGOs may also become important, as is "cell phone journalism."

Some non-conflict, non-protest domains that could be addressed with suitably broad event data

- Illiberal democratic transitions:
  Turkey, Hungary, Poland, Hong Kong, maybe USA
- 21st century right-wing populist movements—Tea Party (USA), Brexit (UK), Yellow vest (France)—vs. 1920s-30s right-wing populism in Italy, Germany, Spain, UK, USA
- Diversionary hypotheses
  - External: Countries take advantage of international crises to escalate in other domains (1956: simultaneous occurrence of Suez crisis and Soviet invasion of Hungary)
  - Internal: Countries escalate in international relations to distract from internal problems (Falklands/Malvinas 1982; Russian annexation of Crimea 2014)

Event data would not be used exclusively here but would be a useful and inexpensive (if already coded) supplement to other measures such as opinion polling and economic data.

Machine learning

New opportunities from machine learning

- ML methods recently have been successful in a number of "artificial intelligence" problems previously thought to be unsolvable. Ask Siri or Alexa if you don't believe this.
- 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

Methods being used in ViEWS forecasting tournament

Contributions: *pgm* (PRIO-GRID month)

1. Event Data and Spatio-Temporal Graph Convolutional Networks
2. 'Dynamic time warping'
3. Hierchical hurdle model and weapons data
4. Peacekeeping data
5. State-space model
6. Automated ML systems
7. Spatial convolutional RNNs
8. Geographical data, XGBoost

Contributions: *cm* (country month)

1. Adaptive elastic net
2. Expert survey
3. 'Dynamic time warping'
4. Ensembles of RF and GBMs
5. Recurrent neural networks
6. Past violence and news sources
7. Automated ML systems
8. Wikipedia statistics
9. Markov models

Risks in machine learning models

- Over-fitting
- It is not clear that political 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

And from nearby in the *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 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
- Genetic algorithms
- Logistic regression, which not infrequently is "embarrassingly effective"
- And...

And the monster in the living room: recurrent neural networks (deep learning)

- These appear to be able to extract pretty much all available signal in a set of data
- They are hugely computationally expensive but now benefit from specialized hardware ("GPU"s) originally developed for accurately rendering splattering zombie brains in video games. Apple's new M1 chip has neural network hardware on the chip.
- Neural networks have always been good at dealing with missing data—which they treat as information—and non-linear relationships
- Some variants are specifically designed to work with sequences, others with two-dimensional data
- Methods of extracting key variables are less developed but exist and can be developed

Weaknesses with respect to conflict forecasting

- We are dealing with rare positive cases (that is, conflicts), even in very large data sets. Conflict outbreaks, as opposed to continuations, are even rarer

- "Trigger sequences" may be a cognitive illusion, so sequence-based methods will not provide advantages over structural methods. Spatial relations, on the other hand, are probably quite real.

- To date, random forests have been the more important breakthrough than neural nets

Open source

Open source, open access, open collaboration

- There is a strong if incomplete norm towards open sharing of data and methods
  - Unintended consequence: PITF "forecasting tournament" cannot be published in a major journal because it used proprietary data—the baseline data has 2,700 variables—that cannot be archived in replication sets. The results are, however, still available on SSRN.
  - The inability to share source texts is clearly a concern in news-report-based datasets such as ICEWS and MID, though URLs can be shared.

- By all available evidence, US government forecasting projects are using similar methodologies to those available in open sources; in fact they are increasing lagging behind Europe on this

- We now have significant NGO and academic work, and an international "epistemic community" has developed around the topic.

Open Source Software

MySQL

python

mozilla Firefox

ubuntu

source forge

T R

perl

Ruby PROGRAMMING Language

WordPress.com

APACHE HTTP SERVER

R

LibreOffice The Document Foundation

GNU run free run GNU

stackoverflow

FREE SOFTWARE FOUNDATION

Open versus proprietary software

I'm not exactly a neutral observer on this issue...

- The open source environment for both natural language processing and event coding is now extraordinarily rich and largely has standardized on the Python programming language. It is thoroughly international.
- Open source software is nonetheless only "free as in puppy:" very substantial investment of labor is required to effectively use a complex open source system
- Continued maintenance and documentation of an open source system depends on the development of a large user community: there are serious network effects in operation
- Many academic incentive structures are not aligned with the production of quality open source systems

Similar issues in. . . astrophysics

Computational astrophysics for the future

Simon Portegies Zwart
+ See all authors and affiliations

Science 07 Sep 2018:
Vol. 361, Issue 6406, pp. 979-980
DOI: 10.1126/science.aau3206

Scientific discovery is mediated by ideas that, after being formulated in hypotheses, can be tested, validated, and quantified before they eventually lead to accepted concepts. Computer-mediated discovery in astrophysics is no exception, but antiquated code that is only intelligible to scientists who were involved in writing it is holding up scientific discovery in the field. A bold initiative is needed to modernize astrophysics code and make it transparent and useful beyond a small group of scientists.

Can't resist sharing this...

dinosource

Astrophysics phrase for poorly documented laboratory software written on the assumption it would only be used for a couple years but still in use, typically endlessly patched, and by multiple projects, two or three decades later.

Issues for astrophysics software relevant to event data used in forecasting models

- Open access to source code is essential for scientific progress and integrity: "secretly developed codes are of no help to the community and produce unverifiable results."
  - Not doing well here: Cline Center, TERRIER and Phoenix are coded with open PETRARCH-2 but the more widely used ICEWS and GDELT use secret coding engines
- Open standards for interchange of program parameters
  - Reasonably okay: ICEWS actor dictionaries are open if odd; TABARI/PETRARCH family is a de facto standard
- Modularized—LEGO blocks—components
  - Doing very well here with modular formatters, parsers, coders, geolocation, pipelines
- Core components need to be available that have been written and documented to industry standards, not laboratory standards
  - Still needs work: PETRARCH family has very poor documentation; TABARI/JABARI and Serif/ACCENT had professional programming, though only TABARI is open

Additional issues

- Many new models have numerous hyperparameters which are often set in rather ad hoc fashions (or, at best, optimized for a specific set of data)
- In rare events analysis, continuity is critical: we need to be able to generate comparable data across many years and many different projects
- There is increasing concern about assessing the implicit biases in models
- We owe transparency to the people who actually are funding us, whether directly through taxes or indirectly through the extortionate oligopoly profits accruing to foundations and philanthropists

Academics: remember that publishing in paywalled journals is equivalent to burying your work beneath a pile of radioactive sludge in Antarctica

The software singularity

- Code for doing almost anything is now available for free and has an effective support community on Stack Overflow: things that once took months now can be done in hours
  - `newspaper3k` downloads, formats and updates news scrapping in 20 lines of Python
  - universal dependency parses provide about 90% of the information required for event coding
  - easily deployed data visualization dashboards are now too numerous to track
- This enables development to be done very rapidly with small decentralized "remote" teams rather than the old model of large programming shops
  - In the software development community in Charlottesville, our CTO group focuses on this as the single greatest current opportunity, and doing it correctly is the single greatest challenge

That other singularity?: no, sentient killer robots are not about to take over the world, and you're going to die someday. Sorry.

CHALLENGES

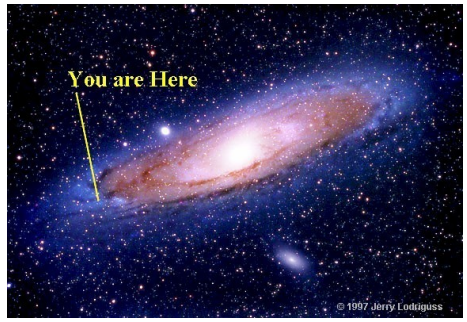
Paradox of political prediction

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 focus on the unexpected (Kahneman).

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.

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, but there are not very many of them.

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 mortgage 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.

Models matter

Arab Spring is an unprecedented product of the new social media

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

Arab Spring is an example of an instability contagion/diffusion process

- Eastern Europe 1989-1991, OECD 1968, CSA 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

Are trigger models simply a cognitive illusion?

- Human experts *assert* they are basing predictions on trigger sequences but is may simply be an artifact of the dominance of episodic associative memory (Kahneman)
- To date, statistical studies have not found that detailed event-based models provide a predictive advantages over structural models at the 6 to 24 month horizon
  - Event data can *substitute* for structural data, so it necessarily contains meaningful information. But it doesn't appear to contain *additional* information.
  - However, this is using traditional aggregated linear time series models: sequence-based methods might do better

Data quality is more important than quantity

- We have apparently passed the "inversion" where more information on the web is computer-generated than human-generated. Some of it is harmless; much of it is not
- The social media platforms clearly will not police themselves: anger, greed, and delusion are their business model
- Various event data issues that need to be seriously addressed
  - False positive rates
  - Duplication rates
  - Urban and other geographical/socioeconomic biases
  - "Media fatigue," particularly in conflict zones
  - Biases in existing event data sets primarily focusing on violent conflict

Making this relevant to the policy community

This is a two-way street.

- The conflict policy community needs to become as sophisticated in evaluating and integrating quantitative models as their counterparts are in economics and public health.
- Academic researchers need to focus on questions and methods relevant to policy and not just "interesting." And/or easy to study. And/or publishable after a five-year lag. And/or accessible only on a publisher's web site for a $40 per article fee.
- Both sides need to work on common standards for evaluating the quantity and robustness of results.
- Both sides need to understand the vocabularies, incentives and cultures of the other.

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?
  - "Policy relevant forecast interval" is around 6 to 24 months
  - USAID/FAO famine forecasting model
  - 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

Final thoughts

400 Years Ago: Francis Bacon establishes principles of modern science

New Method (*Novum Organum*) [1620]

- Scientific method based on the primacy of observation and induction
- Science should be open, in contrast to the secrecy of the alchemists
- Science should benefit society as a whole—also a contrast to the alchemists—and is deserving of state support

- We should not have “one data set or model to rule them all”
- Follow the approach of hurricane and snowstorm forecasters who triangulate results of multiple independently developed data sets and models which have different assumptions and strengths

"What would this look like if it were easy?"—Tim Ferris

Table: Cost to put one person in Earth orbit, 2019 $US-millions

Project	Development	Per-person cost
NASA Space Shuttle	$27,400	$170
NASA Orion (projected)	$23,600	$291
SpaceX CrewDragon	$ 1,700	$ 55

Economist, 4 June 2020

Possibly apochryphal:
Around 2014, Google assessed the performance of deep neural networks combined with word embeddings in machine translation projects, and decided to discard 500,000 lines of hand-crafted code developed over a decade, replacing this with 500 lines of Tensorflow calls.

Thank you

Email:
schrodt735@gmail.com

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

Links to data and software: `http://philipschrodt.org`

Blog: `http://asecondmouse.org`