Technical Political Forecasting: Opportunities and Challenges

Philip A. Schrodt

Parus Analytical Systems schrodt735@gmail.com

Institute for Security Studies Pretoria, South Africa 16 May 2014 What are event data?

WEIS Primary Categories

01	Yield	11	Reject
02	Comment	12	Accuse
03	Consult	13	Protest
04	Approve	14	Deny
05	Promise	15	Demand
06	Grant	16	Warn
07	Reward	17	Threaten
80	Agree	18	Demonstrate
09	Request	19	Reduce Relationship
10	Propose	20	Expel
		21	Seize
		22	Force

News Story Example: Example: 18 December 2007

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

The Turkish attacks in Dohuk Province on Sunday—involving dozens of warplanes and artillery—were the largest known cross-border attack since 2003. They occurred with at least tacit approval from American officials. The Iraqi government, however, said it had not been consulted or informed about the attacks.

Massoud Barzani, leader of the autonomous Kurdish region in the north, condemned the assaults as a violation of Iraqi sovereignty that had undermined months of diplomacy. "These attacks hinder the political efforts exerted to find a peaceful solution based on mutual respect."

New York Times, 18 December 2007

http://www.nytimes.com/2007/12/18/world/middleeast/18iraq.html?_r=1&ref=world&oref=slogin (Accessed 18 December 2007)

TABARI Coding: Lead sentence

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111
Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

TABARI Coding: First event

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111 Source: IRO GO

Target: TUR

Event Code: 223

Source: TUF

TABARI Coding: Actors

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111 Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

TABARI Coding: Agent

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111 Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUI

TABARI Coding: Second event

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111 Source: IRQ GOV

Target: TUR

Event Code: 223

Source: TUR

TABARI Coding: Second event target

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111 Source: IRQ GOV

Target: TUR

Event Code: 223 Source: TUR

Torget: IDOVDD

TABARI Coding: Agent

BAGHDAD. Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111 Source: IRQ GOV

Target: TUR

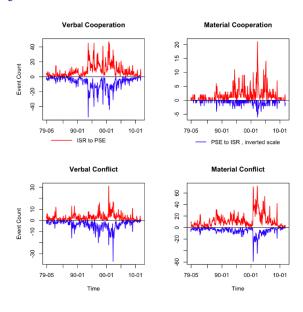
Event Code: 223 Source: TUR

Source: TUR

Quad Counts

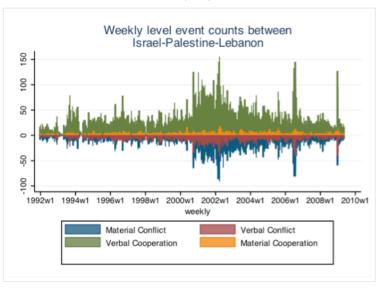
- Verbal Cooperation (VERCP): The occurrence of dialogue-based meetings (i.e. negotiations, peace talks), statements that express a desire to cooperate or appeal for assistance (other than material aid) from other actors. CAMEO categories 01 to 05.
- Material Cooperation (MATCP): Physical acts of collaboration or assistance, including receiving or sending aid, reducing bans and sentencing, etc. CAMEO categories 06 to 09.
- Verbal Conflct (VERCF): A spoken criticism, threat, or accusation, often related to past or future potential acts of material conflct. CAMEO categories 10 to 14.
- ▶ Material Conflict (MATCF): Physical acts of a conflictual nature, including armed attacks, destruction of property, assassination, etc. CAMEO categories 15 to 20.

KEDS Project Levant Data, 1979-2010



KEDS Project Levant Data, 1992-2010

Visualization by Jay Yonamine



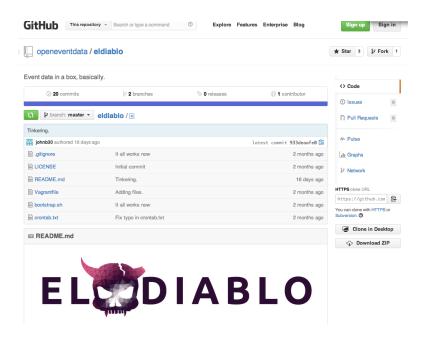
Why Event Data are well suited for predicting political change

- Structural indicators such as GDP, infant mortality, past or adjacent conflict change too slowly
 - ► They nonetheless affect the overall probability
- Social media indicators change too quickly
 - Though US government funders are completely obsessed with this at the moment. Tweet that!
- Newsworthy events are "just right"
 - ▶ And we've got the models to prove it
 - Which is why they are "newsworthy"
 - Structural indicators either are reflected in the patterns of events, or can be additional covariates

EL:DIABLO

Event Location: Dataset in a Box, Linux Option

- ► Full modular open-source pipeline to produce daily event data from web sources
- Scraper from white-list of RSS feeds and web pages
- Event coding from any of several coders: TABARI, PETRARCH, others
- Geolocation: Penn State "GeoVista" project coder, UT/Dallas coder
- Conventional reduplication keeping URLs of all duplicates
- Additional feature detectors are easily added
- Designed for implementation in Linux cloud (e.g. Linode: \$20/month)



PETRARCH

- ▶ Written in Python, in contrast to the C++ TABARI
- ► Full parsing using the Penn Treebank format and Stanford Core NLP. This handles the noun/verb/adjective disambiguation that accounts for much of the size of the TABARI dictionaries
- Synonym sets from WordNet
- ► Identifies actors even if they are not in the dictionaries
- ► Extendible through program "hooks": "issues" facility
- Codes at about 150 sentences per second, about a tenth the speed of TABARI but cluster computing is now readily available
- ► Problem: TABARI dictionaries—based on shallow parsing—do not translate well to the higher precision of full parsing

Open Event Data Alliance

- Institutionalize event data following the model of CRAN and many other decentralized open collaborative research groups: these turn out to be common in most research communities
- ▶ Provide at least one source of daily updates with 24/7/365 data reliability. Ideally, multiple such data sets rather than "one data set to rule them all"
- Establish common standards, formats, and best practices
- ▶ Open source, open collaboration, open access

Extending the event ontologies

CAMEO and IDEA were derived from earlier Cold War event ontologies (WEIS, COPDAB, World Handbook) and consequently miss substantial amounts of political behavior that is currently relevant.

- natural disaster
- disease
- criminal activity
- financial activity
- refugees and related humanitarian issues
- human rights violations
- electoral and parliamentary activity

Technical Political Forecasting: The Debate



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





PRINT | TEXT SIZE . | EMAIL | SINGLE PAGE

Predicting the Future Is Easier Than It

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

▶ ARGUMENT

Looks

Factors encouraging technical political forecasting-1

- 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 effect
 - Demographic and epidemiological forecasting
 - ► Famine forecasting: USAID FEWS model
 - Example: statistical models for mortgage repayment were quite accurate
 - ► Moneyball
- ► Technological imperative
 - Increased processing capacity
 - ► Information available on the web
 - "Moore's Law states that computing power doubles every 18 months. Human cognitive ability is pretty much a constant. This leads to some interesting and not always desirable substitution effects"

Larry Bartels, Princeton University

Factors encouraging technical political forecasting-2

- Demonstrated utility of existing methods, which tend to converge on about 80% accuracy
 - ▶ Political Instability Task Force
 - ICEWS
 - "Big Data" analytical methods
- 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

Like 95 Apr 21st 2012 | from the print edition

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-present
- IARPA ACE and OSI
- Peace Research Center Oslo (PRIO) and Uppsala University UCDP models

Convergent Results

- ▶ 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
- ▶ 80% 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."

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	protests, strikes, government crises
state authority	
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
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 fregime 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

Political Instability Task Force (AJPS 2010)

FORECASTING POLITICAL INSTABILITY

201

Table 2 Out-of-Sample Prediction Exercise for Observed Onsets of Instability, 1995–2004

A. Countries That Had Instability Onsets, 1995–2004. Quintile/decile in model score rankings based on 2-yr. prior data						
Year	Top Decile	Second Decile	Second Quintile	Third Quintile		
1995	Armenia, Comoros	Belarus				
1996	Albania, Niger, Zambia		Nepal			
1997	Cambodia, Congo-Brazz.					
1998	Guinea-Bissau, Lesotho			Serbia/Montenegro		
1999	Ethiopia, Haiti					
2000		Solomon Ils., Guinea*				
2002	Cote d'Ivoire					
2003	Central African Republic					
2004	Iran*	Yemen*		Thailand*		

B. Tabulation of All Country-years, 1995–2004. Model estimates based on censored data, using only sample data from prior to year of forecast (countries w/population over 500,000, no ongoing conflict, at least two years old)

	Countries with Instability in $t+2$	Countries Remaining Stable
Predicted for Instability (Top Quintile)	18	233
Predicted for Stability (Not Top Quintile)	3	992
N = 1,246 Percent Classed Correctly	85.7%	81.0%

Number of instability onsets, 1995–2004: 21. Number of instability onsets in top quintile of model scores: 18 (86%).
*Cases added to the problem set in 2005 update.

This is ca. 2010

^{*}Cases added to the problem set in 2005 update

PITF Results, ca. 2005

TABLE 1 Results of Global Analysis of Onsets of Instability

	Full Problem Set		Civil War Onsets		Adverse Regime Change Onsets	
Independent Variables	Coefficient (S.E.)	Odds Ratio (95% CI)	Coefficient (S.E.)	Odds Ratio (95% CI)	Coefficient (S.E.)	Odds Ratio (95% CI)
Regime Type (Full Autocracy as	Reference)					
Partial Autocracy	1.85***	6.37	1.94***	6.98	2.85***	17.32
	(0.47)	(2.53, 16.02)	(0.62)	(2.05, 23.8)	(0.86)	(3.19, 94.0)
Partial Democracy with	3.61***	36.91	3.35***	28.5	5.06***	157.0
Factionalism	(0.51)	(13.5, 101)	(0.73)	(6.86, 118)	(1.02)	(21.1, 1164)
Partial Democracy without	1.83***	6.22	.981	2.67	2.58***	13.23
Factionalism	(0.54)	(2.17, 17.8)	(0.79)	(0.57, 12.4)	(0.91)	(2.20, 79.5)
Full Democracy	0.981	2.67	.545	1.73	1.26	3.51
	(0.68)	(0.70, 10.2)	(0.92)	(0.29, 10.4)	(1.09)	(0.42, 29.5)
Infant Mortality†	1.59***	6.59	1.64***	4.19	1.38*	4.56
•	(0.35)	(2.91, 14.9)	(0.48)	(1.82, 9.60)	(0.58)	(1.30, 16.0)
Armed Conflict in 4+	3.09***	22.0	2.81***	16.7	.091	1.10
Bordering States	(0.95)	(3.42, 142)	(0.82)	(3.36, 83.0)	(1.49)	(0.06, 20.4)
State-Led Discrimination	0.657*	1.93	1.17***	3.23	502	0.61
	(0.30)	(1.08, 3.45)	(0.36)	(1.59, 6.55)	(0.62)	(0.18, 2.04)
N = Total (Problems, Controls)	468 (117, 351)		260 (65, 195)		196 (49, 147)	
Onsets Correctly Classified	80.3%		80.0%		87.8%	
Controls Correctly Classified	81.8%		81.0%		87.8%	

^{***} p < 0.001, ** p < 0.01, * p < 0.05. †Odds ratios for continuous variables compare cases at the 75th and 25th percentiles.

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

Multi-disciplinary challenges

Big Data: Machine learning researchers routinely use social science data to construct models. Many of these achieve high accuracy in split-sample tests, to the point where these researchers simply assume that things are predictable.

IARPA ACE "Good Judgment Project" (Tetlock): While *most* forecasters do no better than chance, a small number of "super forecasters" perform significantly better than chance over long periods of time and large numbers of questions. Furthermore these individuals exhibit common characteristics and strategies, and to a limited extent, these can be taught. *Economist: The Year 2014*

The Forecaster's Quartet

- Nassem Nicholas Taleb. The Black Swan (most entertaining)
- Daniel Kahneman. Thinking Fast and Slow
 (30 years of research which won Nobel Prize)
- Philip Tetlock. Expert Political Judgment (most directly relevant)
- Nate Silver. The Signal and the Noise (high level of credibility after perfect 2012 electoral vote predictions)

Data!



































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

Open Source Software





























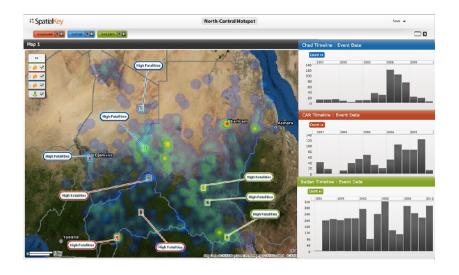




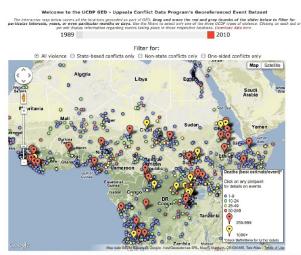
Computationally-intensive methods

- ► 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

Geolocation: ACLED Geospatial



Geolocation: Uppsala Conflict Data Project 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.

Open source, open access, open collaboration

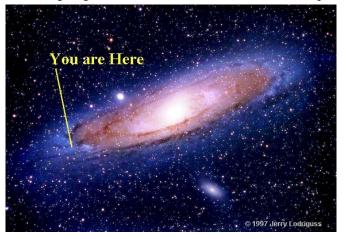
- There is a strong if incomplete norm towards open sharing of data and methods
- By all available evidence, US government forecasting projects are using similar methodologies to those available in open sources; in fact they are probably lagging somewhat behind this
- We now have significant NGO and academic work, and an international "epistemic community" has developed around the topic.

Welcome to the new normal

- Rifkin [NYT March 2014]: The most disruptive technologies in the current environment combine network effects with zero marginal cost
- Key: zero marginal costs: open source software is "free-as-in-puppy"
- Examples
 - Operating systems: Linux
 - ▶ Statistical software: *R*
 - ► Encyclopedia: Wikipedia
 - Commercial photography: Shutterstock (55K photographers; 30K new images per day) vs Getty Images in \$5B/year market

Challenge: Black swans

Ideal forecasting targets are neither too common nor too frequent



Good Judgment Project: look for events with a 10% probability

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.

Heterogeneous environments

- Per Pinker, Goldstein, Mueller, etc, is the system changing significantly while we are trying to model it? How far back are data still relevant?
- ► How different are various types of militarized non-state actors? For example, how much do al-Qaeda and international narcotics networks have in common?
- ▶ Will changes in the technological environment—internet, UAVs, various monitoring technologies—change probabilities?
- ▶ We are also using a more heterogenous set of forecasting methods, and probably do not understand their weak points as well as we understand those of regression-based models.

Note that all of these are complicated by rare events—some of which may be black swans—since it limits the number of observations we have on the dependent variable.

Irreducible sources of uncertainty

- 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%
- ▶ 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."
- Quasi-random structural error: Complex and chaotic deterministic systems behave as if they were random under at least some parameter combinations

Balancing factors which make behavior predictable

- ► Individual preferences and expectations, which tend to change very slowly
- Organizational and bureaucratic rules and norms
- 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

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.

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

Thank you

Email:schrodt735@gmail.com

Slides:

http://eventdata.parusanalytics.com/presentations.h

Software: https://openeventdata.github.io/

Papers:

http://eventdata.parusanalytics.com/papers.html