This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

Predictive Models for Political Instability[1]

White Paper in Response to SBE 2020: Future Research in the Social, Behavioral & Economic Sciences

Philip A. Schrodt
Department of Political Science
Pennsylvania State University
schrodt@psu.edu

Abstract

Over the past decade, two very large-scale U.S. government projects—the multi-agency Political Instability Task Force (PITF) and DARPA's Integrated Conflict Early Warning Systems (ICEWS)—have invested tens of millions of dollars in the development of statistical early warning systems for forecasting a variety of measures of political instability. While both systems have had considerable input from political scientists, there has been little interaction between this work and NSF-funded academic research. PITF and ICEWS have been very successful in demonstrating that we now have sufficient data and appropriate statistical methods to create forecasting models that have substantially higher out-of-sample accuracy than traditional "expert" forecasting, but a substantial number of basic research questions remain unanswered. These include development of ensemble methods for the integration of multiple models, further work on statistical time series and pattern recognition models, assessing causality and counter-factual inference issues, and integration of qualitative assessments into the quantitative models. These basic research issues cross a number of disciplinary and substantive boundaries, would contribute substantially to the intellectual infrastructure in the field, and assist in the transfer of this U.S. contract research to both the academic methodology community and the larger non-governmental organization community.

The Problem

The problem of forecasting political instability has gone through a series of cycles of interest since the development of quantitative methods in political science in the 1960s. Private foundations (Ford, Rockefeller) and the U.S. government invested substantial efforts in this problem in the 1970s, largely without success. Inspired in part by the considerable success of famine forecasting models, another cycle occurred in the 1990s, expressed largely in efforts by international governmental and nongovernmental organizations (Davies and Gurr 1998). Most recently, the unclassified work in the U.S. by the multi-agency Political Instability Task Force (PITF; Goldstone et al 2010) and DARPA's Integrated Conflict Early Warning Systems (ICEWS; O'Brien 2010) have achieved considerable success, with out-of-sample accuracy on the order of 80%. This is a substantial improvement over human "expert" predictions, which in systematic tests have been shown to be barely better than chance. It thus appears that we now have sufficient data and appropriate statistical methods to create good forecasting models, and these are now being operationally deployed in parts of the policy community.

While there has been substantial academic political science input to both PITF and ICEWS, the feedback from these projects to the academic community has been limited. Despite the importance of prediction as one of the defining characteristics of "scientific" research—for example it is the sole characteristic distinguishing orthodox science from formalized pseudo-sciences such as astrology—and the widespread use of forecasting models in several of the social sciences (e.g. economics and demography), such models have not been a priority for most academics in political science.

[1] This work is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/3.0/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

This is a missed opportunity, since the problem of political prediction engages a substantial number of basic research questions which cross a number of disciplinary and substantive boundaries. Work on these issues would contribute substantially to the intellectual infrastructure in the field as a whole, not just to the policy community.

The Implications

A focus on forecasting has two general implications for the direction of NSF-funded work. First, it would create an alternative to the frequentist null hypothesis paradigm that has dominated quantitative political science research for most of the past fifty years. That approach does not consistently yield models with predictive accuracy, and as a consequence, the large-scale projects such as PITF and ICEWS must develop their own basic research to deal with these problems rather than using tools developed in academic basic research. As indicated below, there are a large number of clearly delineated problems and approaches relevant to this issue; what is missing is focused funding and integration of the results.

Second, this work would be clearly consistent with NSF's mandate in PL 81-507, which established the NSF in 1950 by providing fundamental research that supports "programs that provide a source of information for policy formulation."

Provocative Research

The following list indicates some of the core basic research questions relevant to political prediction and some of the current work that is being done.

Robust predictions of rare events

One of the major problems confronting political forecasting models is that the onset and cessation of the behaviors of interest occur very rarely. Naïve autoregressive models—the behavior at time *t* will be whatever it was at time *t-1*—are very accurate, but of little utility for policy (or theoretical) purposes. Frequentist asymptotic methods are similarly of limited utility. There is substantial rare-events work in the quantitative biomedical literature, and this has been successfully applied to forecasting work, notably King and Zeng's critique of PITF's predecessor, the State Failures Project, but more systematic work addressed specifically at the circumstances of political forecasting—where lead times are shorter than in medical forecasting, and indicators tend to be event-based rather than structural—is needed.

Out-of-sample assessment and robust [re-]estimation

Prediction problems are usually evaluated on out-of-sample accuracy and ROC curves, rather than the hoary p-value-and-stars significance tests that dominate most social science journals. The challenge in such modeling is using the existing data to find models that do not "fit the error," as well as the related issue of determining the stability of coefficient estimates over time and knowing when to re-estimate the model. Bootstrapping methods and in-sample techniques such as K-fold cross-validation are two approaches to this, but more work, and more experimentation in the specific domain of political forecasting, would be useful.

Causality

Causality and related issues such as counterfactual analysis is one of the most difficult issues in all of the non-experimental sciences, and the frequentist modeling approach is particularly weak on this. The issue is core to policy-oriented forecasting when possible interventions are contemplated. While dormant for a number of years, causality is getting renewed interest: it was the theme of the 2009 Political Methodology Summer Meeting at Yale, for example. It is being addressed directly by the work of Green, Fearon, Goertz, and Mahoney among others, and also in the closely related issue of inference using case-matching methods, e.g. the work of Sekhon and Imai.

Pattern recognition models

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

Almost all of the forecasting by both PITF and ICEWS is done with conventional frequentist statistical models, usually working out of a maximum-likelihood perspective. There are a variety of alternative algorithmic pattern recognition techniques that might also be appropriate for these problems. Examples include neural networks, variations on Fourier analysis, assorted principal components and latent variable models, hidden Markov models and conditional random fields, a variety of sequential, functional, topological and hierarchical clustering algorithms, and genetic algorithms and simulated annealing methods. Many of these methods are easily implemented with existing open source software (*R* or specialized packages such as *Weka*) but they have been barely explored as possible bases for political instability forecasting models.

Ensemble models

Existing forecasting models are reasonably good at getting the modal cases of political instability—even in rare events situations—but it is likely that predicting outlying cases, which are often those of greatest interest from both theoretical and policy perspectives, will require ensemble methods that combine multiple models that are sensitive to different circumstances. Such methods—for example Bayesian model averaging—are now being used very effectively in domains such as quantitative weather forecasting (e.g. researchers at the University of Washington) and macro-economic forecasting (e.g early work by Granger, Zellner; more recently by Geweke, Clyde). Additional research applying these techniques to the specific circumstances of political instability forecasting would be very useful.

Estimation of hybrid ABM/empirical models

Agent-based models (ABMs) are one of the most popular methods in applied deductive modeling, since these can be used to quickly instantiate very complex sets of assumptions and—in contrast to simpler analytical models such as game theory that dominated earlier approaches—the implications of those behaviors can be quickly evaluated using contemporary computer hardware. However, there is still no clear way of estimating or evaluating these parameter-rich models, nor consistent standards for comparing the multiple outcomes that they produce when they contain stochastic elements. There is a need to develop a hybrid approach that preserves the flexibility of ABMs while still grounding them in the available data.

Integration of SME input

Despite the generally low accuracy of subject matter expert (SME) forecasting, in some circumstances qualitative SME input may be useful as one factor in quantitative models. For example, at times SME assessments may accurately integrate a large number of diffuse indicators (e.g. “political mood”), and SME knowledge might be useful in outlier situations where comparable cases are unavailable in the existing data. In addition, if SME input could be evaluated in conjunction with objective indicators, the relative contributions of traditional and systematic approaches could be assessed. However, at the present time there are no widely accepted approaches for doing this: Bayesian methods would appear to be very promising but to date most Bayesian work in political science has used uninformative priors, and the problem of the best methods of eliciting Bayesian priors from qualitative expertise could use additional research. This is a very general issues and the implications of such work could extend well beyond the domain of political forecasting.

Real-time updating of data sets

True prediction problems require data that are regularly updated. With the low-cost and real-time availability of textual information on political events (e.g. news reports through web sites and RSS feeds) and sentiment (e.g. blogs and social media), much of this can now be done automatically; this approach is being pursued by ICEWS in particular. While some automated coding systems currently exist, these could be substantially enhanced by integrating them with open-source natural language processing systems for tasks such as named-entity recognition/resolution, deep parsing, pronoun coreferencing and machine translation.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

Scientific infrastructure for policy support

Finally, on a somewhat self-interested note, at the present time it is very difficult for those of us who *are* doing this type of work to find assistance from within the academic community. Despite the relatively poor academic job market, I received no qualified applicants for a fairly broadly-defined (and competitively compensated) post-doc I advertised in the spring of 2010. A former colleague now working on the classified side of a technical forecasting project has repeatedly conducted interviews at APSA and MPSA and has found few applicants with relevant skills (though those who do have had very successful government careers). While the lead researchers on PITF and ICEWS are Ph.D.s in political science, most of the technical support staff are not, and this leads to considerable inefficiency when staff need to be trained in basic issues such as political science concepts and measurement methods. Further NSF funding for prediction projects in general, and perhaps the sponsorship of some summer workshops on the model of recent workshops on experimental methods and textual analysis might increase the pool and provide a means of communicating current methods, approaches, and best practices to graduate students who appear unlikely to receive such training in depth in their home departments.

References

Davies, J. L. and T. R. Gurr. (Eds.). 1998. *Preventive Measures: Building Risk Assessment and Crisis Early Warning*. Lanham, MD: Rowman and Littlefield.

Goldstone, Jack A., Robert Bates, David L. Epstein, Ted Robert Gurr, Michael Lustik, Monty G. Marshall, Jay Ulfelder, and Mark Woodward. 2010. A Global Model for Forecasting Political Instability. *American Journal of Political Science* 54, 1: 190-208.

O'Brien, Sean. 2010. Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research. *International Studies Review* 12,1: 87-104