

The Identification of Sequential Patterns Preceding the Occurrence of Political Events of Interest*

Kurt Rohloff

BBN Technologies
10 Moulton St.
Cambridge, MA 02138
krohloff@bbn.com

Victor Asal

University at Albany, SUNY
135 Western Ave.
Albany NY 12203
vasal@email.albany.edu

Abstract

In this paper we present an approach to identifying patterns of behavior preceding political instability events in countries from sampled factor data. This process is based on the concept of state-space back-chaining. A list of sampled, quantized factor data sampled over a range of discrete times define a "state-space" of a country, and the list of quantized factor data associated with a country at a particular instance in time defines a state. The state of a country changes over time and instances of two countries changing in the same manner define a pattern. At some country states political instability events of interest (such as the onset of regime change, insurgency, ethnic violence, etc) may be observed to occur. We discuss a method to identify the set of factors which define a state space and pattern for combinations of selected political instability events of interest such as rebellion, insurgency, civil war, etc. This approach can be used to identify the observable behaviors before the occurrence of events of interest. The backwards chaining methodology is implemented in the Java programming language and run over a set of factor data. We present a pattern for domestic political crisis found using this method.

Introduction

A major challenge in computational political science is the problem of identifying symptomatic precursors to political instability events such as rebellion, insurgency, civil war, etc. In this paper we present the "backwards chaining" methodology to discover quantitative changes in the properties of a country which precede political instability events and are exhibited in multiple countries. With this knowledge we would not only be able to better understand the factors that drive (or at least precede) political instability, but we would have an approach to forecasting the occurrence of political instability.

Our approach to pattern discovery is based on the supposition that the phenomena which cause (or at least are related to) the occurrences of Political Events of Interest (EoIs) exhibit similar symptomatic behaviors across multiple EoI occurrences. For example, for countries with rebellions driven

by the desire for freedom by internal ethnic groups commonly exhibit increasing ethnic tension and violence before the occurrence of ethnic rebellions. Our methodology to discovering patterns is served by the use of regularly sampled factor data. A sampled factor is a quantifiable measurement that may be taken (or sampled) from a country at discrete, regular points in time. Example factors potentially include GDP, the rates of occurrence of various words in the national press, the average caloric intake, etc... The sampling period of factor data may be over any regular period (such as yearly, monthly, weekly) as long as the sampling period and measurement methodology is constant. Naturally, some factors may be easier to measure accurately than others.

To identify patterns of changes in factors preceding political instability events we used the backwards chaining methodology. The backwards chaining methodology permits us to identify which factors change identically over a fixed number of time steps in the time period leading up to the occurrences of events of interest in selected countries. The identification of how specific factors change over time leading to the occurrence of an EoI defines a pattern in the context of our backwards chaining methodology.

In the context of our backwards chaining methodology the set of factors represents the condition of a country at particular moments in time. We use the equivalence relationships for the various factors to identify what possible discrete combinations of factors are reachable by the countries. We combine multiple factors identified by the backwards chaining methodology to define the multi-dimensional "state" of a country where each factor is a dimension in the state of the country. The set of all possible states for all countries with respect to a set of factors defines a "state space." Based on discovered patterns of changes in factors leading to the occurrence of EoIs, we can generate early-warning forecasts of EoIs if early portions of the patterns are observed in real-time for a specific country. For our experimentation we search for patterns that lead to EoIs driven by (or at least related to preceding) government policy and immediate antecedent behavior. Examples of this preceding (and possibly driving) behavior include shifts in government policy and economic performance. These antecedent conditions can create agitation and spark violence amongst the country's population when expectations are let down or there is a spike in repression. Importantly, contextual information is

*Supported by AFRL/HECS contract FA8650-07-C-7748
Copyright © 2008, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

non-trivial: some factors in a pattern state space might not change over time, but they set an important context for the country's state evolution.

It is also important to note the importance of equifinality in this analysis as well. A country may or may not follow multiple patterns leading to EoI occurrence simultaneously, and there may be multiple means to a same end. For example, a country such as India may contain multiple types of rebellion or may exhibit the antecedents for several rebellions simultaneously.

This section of the report is organized as follows. Section discusses the theoretical basis for our analysis. Section presents our backwards chaining methodology. Section presents and discusses a pattern discovered with our methodology and several patterns for insurgency.

A Theoretical Basis for the Pattern Concept

We start our investigation with the belief that the outbreak of violence will be characterized by equifinality, "many alternative causal paths to the same outcome (George and Bennett 2005)," in other words there is not one cause of EOI outbreaks. There may be a set of factors that make those outbreaks more or less likely but we believe there are a set of potential causal pathways that are likely to lead to outbreak of non-state actor violence towards the state. We are interested in exploring the combinational power of various factors as they lead to EOI's. To do this we build on efforts to use Boolean analysis (Ragin 2000) to understand political activity "while allowing for "multiple causal mechanisms (Chan 2003). Most of this work though starts with collecting data on what is trying to be explained rather than collecting data on both the general environment (the dogs that don't bark - and the cats that never would) as well as those occasions where there is an outbreak of an EOI.

For the backwards chaining process we identify which factors change in the same manner for several time steps leading up to all occurrences of a particular EoI. The combination of all of the factors which all change in the same manner for a fixed number of time steps leading up to all occurrences of a particular EoI define a pattern. By identifying which specific factors which in combination exhibit symptomatic behavior leading to EoI occurrences, we are closer to our ongoing goal of obtaining early warning derivative results informed by factor combinatorics in a semi-automated manner. Our hypothesis is that by identifying such a pattern, one can use this pattern to detect conditions which are precursors to the occurrences of EoIs and hence have a forecasting capability for EoI occurrence.

It is important to note that a country may or may not follow multiple patterns leading to EoI occurrence simultaneously, and there may be multiple means to a same end. For example, a country such as India may contain multiple types of rebellion or may exhibit the antecedents for several rebellions simultaneously.

In our effort to build on Boolean thinking we use the backward chaining approach to go beyond the small n efforts that have been the focus of most Boolean analysis and to extend the effort temporally such that we look for patterns across

variables and across time. From a policy perspective our efforts allow us to identify combinations of changing patterns of behaviors within particular larger sociopolitical contexts that are likely, based on past experience to lead to EOI outbreaks at a temporal point distant from the actual outbreak.

Our analytical effort is being driven by an approach that does not privilege any particular social science theoretical bias related to violent conflict (for example the greed - grievance argument). Instead we have endeavored to draw from the best theoretical models we have been able to find and to operationalize theories in an analytically useful way. Our efforts are directed at creating a synthesis that draws the most successful components from each of the three perspectives into a coherent whole that can be used to understand and predict EoIs. To start with we drew explicitly from the greed (Collier et al. 2003), grievance (Gurr 2000), resource mobilization (McAdam, McCarthy, and Zald 1996), political opportunity structure, (external and demographic) pressures (Tarrow 2001), culture/values (Goldstone 2001) and leadership literatures (Herman and Herman 1989). In this effort we are looking to identify factors that help explain rare events - the outbreak of different kinds of political violence. In general we view EoI emergence, as a product of interactions between causal factors at different levels of a social environment. Some aspects of that social environment are more changeable and stochastic while other aspects are relatively rigid and predictable. These variables combine in different patterns to produce future behavior-in our case patterns of violence.

In our current analysis we focus on the variables that change leading up to the EOI's but we should recognize that the backward chaining approach we are using allows us to also identify the ongoing unchanging factors that go into creating a state space where the changes in particular behavioral and policy factors move a country towards experiencing and EOI. The larger state space of countries primed for an EOI are defined by elements like general political instability married to anocratic (regimes between democracy and autocracy) political structures or low levels of militarization. What we find is that when these types of conditions exist the stage is set for an EOI. The stage though is not what sparks the EOI.

Across the different EOI's we are finding is that a general decline in good expressions and behavior and a rise in bad expressions and behaviors - although not as often as we see the good going down. Certain variables repeat across EOIs as leading behavioral indicators. Specifically:

- The general count of good behavior goes down.
- Efforts at public diplomacy go down.
- Protest behavior tends to go down.

The first two changes are fairly intuitive - good behavior is an encoding of how often "good" expressions appeared in the popular press for a given country over a 1-month time span. The changes latter may be seen as counterintuitive in that we can think of protests as a step forward on the continuum of contentious behavior (McAdam, Tarrow, and Tilly 2001). On the other hand if we envision contention as a

choice between several kinds of contention (each one exacting a cost) then the withdrawal from a non-violent contention may be a sign that opposition groups are repositioning their resources for more violent approaches and governments are expending resources to drive this behavior down - perhaps unwittingly pushing opponents into activities that are much more dangerous. Overall, our approach allows us to model different patterns of change within a broader unchanging state space that leads to violence.

The Backwards Chaining Methodology

In this section we present our methodology for automatically discovering patterns based on the backwards chaining methodology process. We implemented this process in the Java programming language to automatically obtain data from a data server, search over the data in an automated manner to identify key factor changes that precede selected EoIs to identify patterns.

To discover patterns using the backwards chaining methodology we developed algorithms and wrote software to identify factors that change "identically" over a fixed number of sample times in the time period leading up to the occurrences of user-selected EoI advents. We define an equivalence relationship for the factor values based on quantization levels of those factors that was implemented in our factor identification tool. We use that equivalence relationship to determine when changes in factors are similar enough to be called "identical".

For the backwards chaining methodology, we define a pattern for the advent of an EoI to be:

1. A set of factors, and
2. A description on how each of those factors change quantitatively a fixed number of time steps before the advent of an Event of Interest in at least two distinct instances.

The set of factors which define a pattern may include a factor that represents previous occurrences of the advent of the EoI itself.

An example of a hypothetical identification of two factors that change identically in the time preceding the occurrence of an EoI is seen in Figure 1. This figure shows the values of two factors (quality of government and level of corruption) for two countries for several quarters preceding the occurrence of the EoI rebellion. The trajectory of one country is shown using a black line and the trajectory of a second country is shown using a light blue line. In this example, the values of the Quality of Government and Corruption are nearly the same for up to three quarters before the occurrence of Rebellion. Although in all of the pattern examples we discuss in this paper are derived from behaviors in two different countries at two different times, our patterns could be derived from the behavior of two different countries at the same time or even the same country at two different times.

In the context of our backwards chaining methodology, the set of factors that define a pattern represents the specific aspects of the condition of a country at particular moments in time. Based on discovered patterns of changes in factors leading to the advent of EoIs, we can generate early-warning forecasts of EoIs if early portions of the patterns are

observed in real-time for a specific country. Although we are discovering patterns using training factor data inside the AOR, the patterns can be used to forecast EoI occurrences outside of the AOR.

In general, we search for patterns that lead to EoIs driven by (or at least related to preceding) government policy and immediate antecedent behavior. Examples of this preceding (and possibly driving) behavior include shifts in government policy and economic performance. These antecedent conditions can create agitation and spark violence amongst the country's population when expectations are let down or there is a spike in repression. It is important to note that we are searching for and forecasting on factors acting in combination and over time which cause the advent of events of interest. Contextual information is non-trivial: some factors in a pattern state space might not change over time, but they set an important context for the country's state evolution.

An overview of our process of identifying factors for a pattern are as follows: (We describe these steps in detail below.)

1. Identify EoI occurrences for which patterns should be identified.
2. Quantize Factor Data.
3. Determine which factors are identical for all instances for a user-specified number of times steps before EoI occurrence.

This process is presented graphically in Figure 2.

We now discuss this process in more detail:

Identify EoI advents for which patterns should be identified.

This step is used to identify the primary operational input to the backwards chaining process in addition to the factor data. We use our backwards chaining approach for sequential pattern identification with the understanding that not all EoI advents of the same type are driven by the same underlining

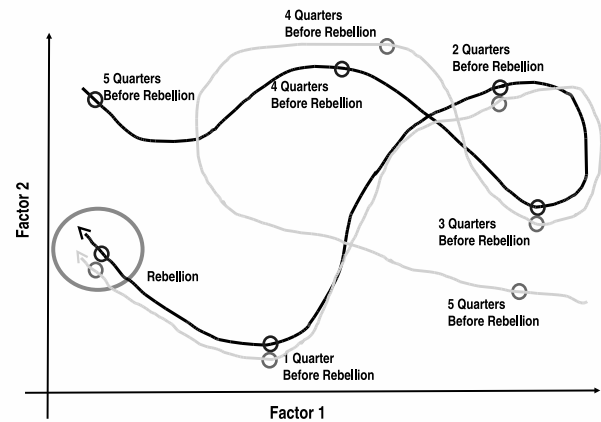


Figure 1: The identification of factors leading to an EoI.

process. (For example, international crises may be driven by ethnic divisions, resource contention or any other number of driving forces.) This step is primarily a user-driven process to select countries and times where interesting patterns are most likely to be found. One of our main hypotheses is that if we attempted to identify patterns using unrelated EoI ad-vents as input, then we would not find meaningful patterns, if we were to find any patterns at all.

Quantize Factor Data.

In order to process the factor data to identify what factors change identically for a number of time steps preceding an EoI, from set of raw factor data for each country, we create two quantized factors. One quantized factor is a straight linear quantization of the raw factor data, and a second quantized factor is a linear quantization of the natural logarithm of the raw factor data. The process of creating two sets of quantized factor data from one raw factor is shown in Figure 3.

Although the straight linear quantization factors provide an indication of the relative value of a factor changes over time, this approach cannot easily account for order-of-magnitude variations in factor values from country to country or from time to time. We use the logarithmic quantization because provides an indication of how the order of magnitude of a factor value changes over time.

To generate the logarithmic factor data from the source factor data we used the natural logarithm function in Java. When factors contain data that was undefined for the natural logarithm operation (such as when the factor data was less than or equal to zero), we did not use the logarithmic quantization of that factor.

For our initial experiments we obtained multiple meaningful results with factor data that was automatically quantized to either 3, 4 and 5 quantization levels for our experimentation factor data. Note that we use all of the factor data to perform the quantization operation to ensure that we have a sufficiently broad view of how the various factors change

over time. If we were to quantize the factor data using only the factor values that define the EoIs, the factor quantizations would be skewed to those countries.

Determine which factors are identical for all instances for a user-specified number of times steps before EoI occurrence.

For each EoI advent selected by the user to construct the pattern we identify on an automated quantized factor by quantized factor basis which of the factors have identical values for a user-defined number of time steps before the advent of the EoI in the countries where the EoIs occurred. These collections of factors and how they change over time leading up to the occurrence of an EoI define a pattern.

We list the patterns we found in a later section and on our web site, but in the next subsection we present and discuss a representative pattern we found through our backwards chaining process.

As a preliminary validation of the factors that we found, we tested to see if any countries in the test data AOR have identical historical values such that they match exactly the factors in our patterns. We implemented this functionality in the Java programming language to take discovered patterns and automatically test if any countries in the test data over any date range that we have factor data for matches the pattern. If any country matches the factor values, but there was no EoI present in the country for the period immediately after the pattern match, then we declared that the pattern had a "false alarm". In general we found that there were very few or no false alarms in our discovered patterns. In the discovered patterns that were output by our automated pattern discovery process there is generally a large number of factors which are identical in the time leading up to the advent of EoIs, but a relatively small number of factors which change over time preceding the EoIs. We hypothesized that the changing factors are symptomatic predecessors of the EoIs and the constant factors are either contextual information for the changing factors to precede the advent of the EoI

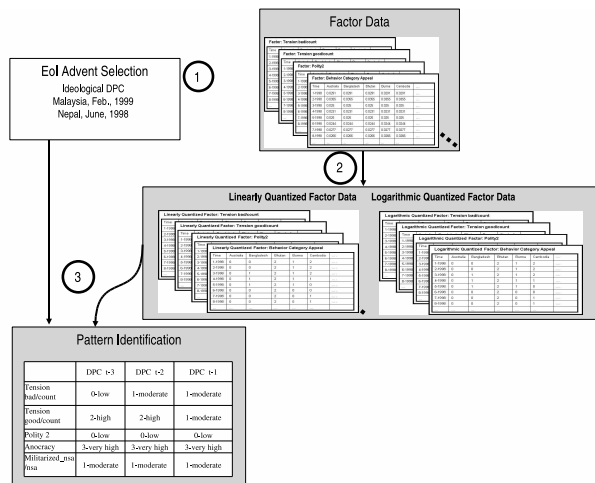


Figure 2: The Pattern Identification Concept of Operations.

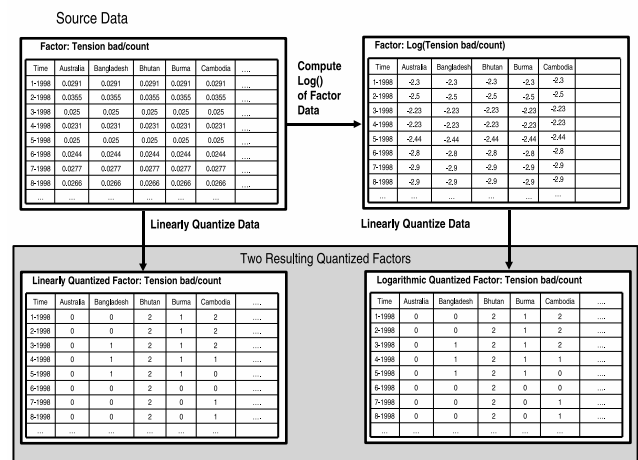


Figure 3: The Data Quantization Process.

or they are extraneous information. To identify the necessary constant contextual factors in our patterns and remove the extraneous constant factors in our patterns, we repeatedly and randomly removed constant factors from the pattern that does not decrease the false alarm rate. We repeated this process many times to find an approximation to the minimal number of constant factors needed to define a pattern for the defining EoIs with a low false alarm rate.

Patterns Discovered Using Operational Data

We now discuss representative results of our pattern search and present several patterns. For our factor data we used a combination of data from the QoG compilation imputed monthly and additional in-house factor data derived from counting the occurrences of certain words in news-feeds. When analyzing the quantized factors selected through the back-chaining methodology to define a pattern, we generally see that a decline of the good expressions and behaviors are leading many of the advents and the rise of bad expressions in the public press and behaviors are as well- although not as often as we see the good going down. Our use of the terminology "low, moderate, high", etc is used to indicate the relative quantization level of the factor at the various time steps. This information is in a sense redundant because we use numeric indications of quantization level (0, 1, 2,), but we included the qualitative indicators because it may not always be clear what the numeric levels represent due to the varying quantization levels across patterns.

Our first example (which can be seen below in Table 1) is one that looks at the advent of Domestic Political Crisis (DPC) using a three-level quantization. It is based on advents of DPC in Malaysia in Jan, 1999 and Nepal in May, 1998. The general context is a moderate (but not low) level of militarized non-state actors as a function of the total number of non-state actors (NSA) within the framework of an autocratic country. As we mentioned above (and this DPC result is a classic example of this) what we are seeing is the rise of bad expressions and behavior as captured in our count or bad "words and actions" and a decline in a similar measure for good "words and actions."

Turning our attention to other patterns we found for insurgencies (Tables 2 thru 4) we offer examples that are at different quantization levels and based on different combinations of cases that tell related but different stories about when violence is likely to break out. In the Table 2 pattern we see a clear example of the fall of good "words" combined with problematic contexts (like high instability combined with very low militarization) can lead to wide scale organized violence. Using a different quantization but building on the same countries we see a different twist to the story. Protests fall from high to moderate. This may be an indication of a switch of strategies on the part of opposition forces as the reposition their resources to more violent means. Note that this still in the context of low or dropping (depending on how measured) counts of general good "words and actions." The pattern in Table 3 uses one different country and several different variables tells a similar story about falling good action and statements while the pattern in Table 4 again identifies a fall in protest as a clear danger sign given the right

contextual variables.

References

- Chan, S. 2003. Explaining war termination: A boolean analysis of causes. *Journal of Peace Research* 40(1):49–66.
- Collier, P.; Elliott, L.; Hegre, H.; Hoeffler, A.; Reynal-Querrol, M.; and Sambanis, N. 2003. *Breaking the Conflict Trap: Civil War and Development Policy*. Oxford: World Bank and Oxford University Press.
- George, A. L., and Bennett, A. 2005. *Case Studies and Theory Development in the Social Sciences*. Boston: The MIT Press.
- Goldstone, J. A. 2001. Toward a fourth generation of revolutionary theory. *Annual Review of Political Science* 4(1):139–187.
- Gurr, T. R. 2000. *People vs. States*. Washington D.C.: United States Institute of Peace.
- Herman, M. G., and Herman, C. F. 1989. Who makes foreign policy decisions and how: An empirical inquiry. *International Studies Quarterly* 33(4):361–387.
- McAdam, D.; McCarthy, J. D.; and Zald, M. N. 1996. *Comparative perspectives on social movements : political opportunities, mobilizing structures, and cultural framings*. Cambridge, New York: Cambridge University Press.
- McAdam, D.; Tarrow, S.; and Tilly, C. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.
- Ragin, C. C. 2000. *Fuzzy-set social science*. University of Chicago Press.
- Tarrow, S. 2001. Transnational politics: Contention and institutions in international politics. *Annual Review of Political Science* 4(1):1–20.

Table 1: DPC for Malaysia in Jan, 1999 and Nepal in May, 1998 Quantization Level 3

Factor	Quantization Method	DPC onset t-3	DPC onset t-2	DPC onset t-1
Tension bad/count	Linear	0-low	1-moderate	1-moderate
Tension good/count	Linear	2-high	2-high	1-moderate
Polity 2	Linear	0-low	0-low	0-low
Militarized_nsa/nsa	Linear	1-moderate	1-moderate	1-moderate

Table 2: Insurgency for Indonesia Jan., 2001, Malaysia August 2001 Quantization Level 3

Factor	Quantization Method	Insurgency onset t-3	Insurgency onset t-2	Insurgency onset t-1
Tension good	Log	2-high	1-moderate	1-moderate
Public statements	Log	2-high	1-moderate	1-moderate
Tension goods/count	Linear	0-low	0-low	0-low
Political globalization	Linear	2-high	2-high	2-high
Good/Tokens	Linear	0-low	0-low	0-low
milperpop	Linear	0-low	0-low	0-low
Log Polcon	Log	2-high	2-high	2-high
Instabl	Linear	2-high	2-high	2-high

Table 3: Insurgency for Indonesia Jan., 2001, Malaysia August 2001 Quantization Level 5

Factor	Quantization Method	Insurgency onset t-3	Insurgency onset t-2	Insurgency onset t-1
Protest	Log	3-high	2-moderate	2-moderate
Anti-market regulatory quality	Log	4-very high	4-very high	3-high
Tension goods/count	Linear	1-low	1-low	1-low
Log investigate	Log	2-moderate	2-moderate	2-moderate

Table 4: Insurgency for Indonesia May., 2003, Thailand Jan., 2004 Quantization Level 3

Factor	Quantization Method	Insurgency onset t-3	Insurgency onset t-2	Insurgency onset t-1
Tension good	Log	2-high	1-moderate	1-moderate
Public Statements	Linear	2-high	1-moderate	1-moderate
Diplomatic cooperation	Linear	2-high	1-moderate	1-moderate
Political Constraints	Linear	1-moderate	1-moderate	1-moderate
Percent of neighbors democratic	Linear	0-low	0-low	0-low
Religious fractionalization	Linear	0-low	0-low	0-low
Political Rights	Linear	1-moderate	1-moderate	1-moderate

Table 5: Insurgency for Laos Feb., 2003, Thailand Jan., 2004 Quantization Level 3

Factor	Quantization Method	Insurgency onset t-3	Insurgency onset t-2	Insurgency onset t-1
Protest Behaviors	Log	1-moderate	1-moderate	0-low
Distrust/bad	Log	2-high	1-moderate	1-moderate
bad/tokens	Linear	0-low	0-low	0-low
Linguistic Fractionalization	Linear	2-high	2-high	2-high
Physical Integrity Rights	Linear	1-moderate	1-moderate	1-moderate