Correlates of State Failure

Pamela Surko and Alan N. Unger

Science Applications International Corporation
16701 W. Bernardo Dr.
San Diego CA 92127
Surkop@saic.com Ungera@saic.com

In response to a request from a senior Clinton Administration official, the CIA established the State Failure Task Force in 1994, a group of researchers from academe, nonprofit institutions, non-intelligence-agency government, and industry tasked to evaluate unclassified historical data to search for correlates of governmental instability. The study amassed data by country by year from 1955 onward for hundreds of demographic, societal, economic, political and environmental indicators, along with an expert assessment for each country-year of whether the country was stable, or had suffered an instability identified as regime crisis, ethnic war, genocide, or revolution. The Task Force has done many analyses; the original tasking was to identify correlates of failure and to attempt to predict globally, two years before onset, which countries would suffer a state failure. Some results of various studies may be found in [ES98] and [ES99].

The resultant data set for all studies is highly imbalanced since thankfully only a few governmental collapses occur each year, out of roughly 160 countries over the threshold population of 500,000 persons used in the study.

The first set of studies to be performed were cross sectional analyses that sought models that would predict the dichotomous variable "failure/nonfailure" for a country year, based on data for that country two years previous. Since only yearly data is used, the two year lag not only provides a model with from 1-2 years predictive lead time (depending on when during the calendar year the crisis occurred), but also insures that data relating to a crisis already in progress is not inadvertently included in the yearly value.

In the 43-year period now under study, there were 135 problem cases identified by subject matter experts. Some of these cases were multi-faceted, containing events of several types, or several events of the same type in rapid succession. Individual events such as revolutionary wars and genocide were clustered into cases if less than 5 years elapsed between events, and thus were classified as one sustained period of instability.

In the same 43 year period, there were roughly 6400 country-years where the country in question was stable, yielding an imbalance in the dataset.

There were several reasons to sample the control (nonfailure) set. The first was the imbalance in the dataset. The second was that, although the cross sectional analysis was not considering the time sequence of country years explicitly, nonetheless the data series are time series, and thus for example the data for Somalia 1982 is highly correlated with the data for Somalia 1983. Since by definition the failure cases are at least five years apart, and because there are so few of them over the 40+ years, they are much less strongly correlated. We wished to balance the correlation in the majority and minority classes. The third was the wish to add some variables to the study that had to be assessed by subject matter experts, coded, and entered by hand, a task too onerous if the entire control set were to be used Therefore for the cross sectional studies, we chose to sample the control set, randomly choosing three control country-years for each failure country-year, matching the controls by year against the failures, to account for world-wide changes in some variables (such as infant mortality) over the 40+ years. The choice of three controls per failure was somewhat arbitrary, but was done to increase the statistical power of the control set while limiting the imbalance, and also to limit the number of data elements that had to be determined and entered by hand.

Several control sets were randomly selected and used in identical fitting procedures; the model results were stable with respect to choice of control set.

Two cross sectional analyses were run for many of the global and regional studies: logistic regression and neural network analysis.

Candidate predictors for the multiple logistic regression model were determined by screening the independent variables with univariate t-tests, to identify those that discriminated cases from controls. The variables in the final models were selected from the candidates using a stepwise regression approach, and the coefficients estimated using maximum likelihood. The fitted model

was then used to classify cases by comparing the regression-generated failure probability to a cutpoint between 0 and 1. If the probability exceeds the cutpoint, the case is a predicted failure. Setting the cutpoint at approximately .25 equalizes sensitivity and specificity when there are three controls per failure cases, as originally targeted.

A multilayer perceptron with one hidden layer and two outputs (failure, nonfailure) was also trained on the same data set. In order to achieve reasonable accuracy on the minority class, the training set consisted of 3 copies of each failure record and one copy of each control record, randomly ordered. This was done rather than down-sizing the control cases further in order to preserve the statistical power of the control set. Random Gaussianly-distributed noise of 2% was added each time an exemplar was presented during training.

It was also found that a slow learning rate was necessary to achieve good accuracy on the minority class. We used a learning rate only .01 the default rate we usually use as a starting value. This lower learning rate gave an apparently poorer accuracy for both classes on the training set, but yielded better generalization and thus better accuracy on the reserved test sets. We also noted that longer training increased the accuracy on the minority class. We were careful to avoid overfitting by being parsimonious with the total number of nodes in the network.

Both logistic regression and neural network models can be tuned to trade off specificity against sensitivity. This tradeoff is usually displayed on a Receiver Operating Characteristic (ROC) curve. The relative cost of false alarms vs. the cost of missing a failure has not been assessed since the models produced are not sufficiently accurate to be used for prediction in practice. Therefore we arbitrarily quote accuracy values for the intersection of the ROC curve with the minor diagonal, that is, where the probability of false alarm is equal to (1 – probability of detection).

This adjustment is made in the logistic regression model by adjusting the cutpoint in the failure probability. In the neural network model, it is made by adjusting the cutpoint on the difference between the two output node activations.

One of the main difficulties we see with analyses such as these, using unbalanced data sets, is the danger of oversearching and overfitting. The statistical power of the data set is limited by the small number of events in the minority class, and this must limit the number of candidate variables to be allowed in predictive models, and must also affect the type of searches we allow through a database with hundreds of candidate variables and only 135 minority-class data points.

Although the work of the Task Force was funded by the CIA's Directorate of Intelligence, neither the Task Force's analyses nor the contents of this report are based on intelligence reporting. The report does not represent the official view of the U.S. Government, the U.S. intelligence community, or the Central Intelligence Agency, but rather the views of the individual authors themselves.

References

[ES98] Esty, Daniel C., Jack A. Goldstone, Ted Robert Gur, Barbara Harff, Pamela T. Surko, Alan N. Unger and Robert S. Chen. The state failure project: early warning research for U.S. foreign policy planning. In Davies, John L., and Ted Robert Gurr, eds. Preventive measures: building risk assessment and crisis early warning systems. 1998, Lanham, MD, Rowman and Littlefield, 27-38.

[ES99] Esty, Daniel C., Jack A. Goldstone, Ted Robert Gur, Barbara Harff, Marc Levy, Geoffrey D. Dabelko, Pamela T. Surko, and Alan N. Unger, State failure task force report: Phase II findings. In *Environmental Change and Security Project Report*, **5**:49-72 (Summer 1999).