Automated Production of Political Indicators

Matt Dickenson
Department of Political Science
Duke University
Durham, NC 27708
`mcd31@duke.edu`

Abstract

Can classification methods help to automate the production of political indicators in near real time? The Militarized Interstate Disputes (MID) dataset, produced by the Correlates of War project, has been widely used in political research and policy discussions over the past three decades. Despite its value for understanding conflict, MID data coding is performed in iterative batches by human coders that lag behind the present by several years. However, reliance solely on human coders is neither necessary nor desirable. This project is the first stage in creating a pipeline to approximate the MID dataset using classification trees and daily event data (GDELT) at a substantial reduction in cost.

1 Introduction

1.1 Motivation

The Militarized Interstate Disputes (MID) dataset has been widely used in political research over the past three decades and is increasingly used in policy applications.[1] Production of MID data is performed in iterative batches by human coders that lag behind the present by several years. For example, the most recent version was released in 2004 and contains data through 2001. An update through 2010 was expected last summer (2012) but is delayed indefinitely. However, reliance solely on human coders is neither necessary nor desirable. Using automated classification methods to classify real-time event data, this project project attempts to approximate the MID dataset while reducing both the labor and financial costs required.

1.2 Related Work

As one of the most widely used dependent variables in international conflict studies, much effort has been devoted to estimating models of MID onset and duration. However, most of this work omits cross-validation and relies on features measured at the annual level, making near real-time (NRT) classification infeasible [1]. A recent shift toward event data with frequent updates (often at the daily level) has helped to address this shortcoming [2, 3, 4, 5, 6].

With this transition toward event data, the political forecasting community has become attune to new challenges and has responded with several established practices. Coding the sentiment of interactions can now be done in NRT using the Tabari system, which aggregates and deduplicates news reports [7, 8]. Event classification can be performed using CAMEO, which uses a schema of twenty event types consisting of material/verbal and cooperative/conflictual actions (see Table 1). These event classifications provide a principled, automated method for exhaustively categorizing the types of events that may consitute an interstate dispute [9].

[Table 1 about here.]

---

[1]The article introducing the most recent version of the MID data has been cited over 500 times on Google Scholar.

Forecasters have also dealt with challenges when aggregating event data up to various temporal levels. Although there is no single best practice, monthly aggregation has become a common strategy [10, 11] and is used in this project. CAMEO event counts are amenable to this approach. They also allow for aggregation up to longer temporal periods than the daily level [3]. Modfiying the features by transforming the raw counts into month-to-month changes (i.e. first-differencing) can also help to simplify the feature set [12].

Interpretability is an important concern in this project due to the policy-relevant nature of the problem and the (potential) need to compare the resulting model to the process used by human coders involved in creating the MID dataset [9]. For this reason, "black box" methods such as Support Vector Machines were judged to be inappropriate. Classification trees (and their continuous counterpart, regression trees, collectively known as CART) offer a nice alternative that is more flexible than GLMs and more interpretable than Random Forests (these two methods should provide lower and upper bounds, respectively, on CART) [13]. CART has been used for event data within conflict studies, and in other fields where researchers encounter similar issues of unbalanced and missing data, such as public health research [14, 15, 16].

2 Model and Methods

2.1 Problem Definition and Data Sources

The problem that this project attempts to solve is the classification of country dyad months (e.g. USA-China-2012-May) as either in conflict or not. To achieve this, we will use real time (daily) event data from the Global Database of Events, Language, and Tone (GDELT), aggregated up to the dyad month level for 1992-present [6]. To measure the dependent variable of conflict, the Militarized Interstate Disputes (MID) dataset will be split into subsets for training and validation [9]. The goal of this project is to replicate and extend MID data coding as accurately as possible using automated procedures. If a reliable method can be developed to replicate the MID data up to 2001, it can then be extended to generate data for interstate disputes since 2001.

2.2 Features of the Data

In work on this project thus far, several important features of the GDELT data have been identified. All events in GDELT are classified according to the CAMEO coding scheme [3]. Within this scheme, there are two major distinctions along two dimensions: acts can be material or verbal, and interactions can be cooperative or conflictual. These four categories provide a rough characterization of how two countries interact within a given period of time. More fine-grain classification, into twenty subcategories, is also provided. Examples of these categories are presented in Table 1.

During the process of aggregating GDELT records into dyad months, the absolute number of events within each of the four major and twenty minor categories was counted. Some example time series of dyads with and without conflict are presented in Figure 1. From these raw counts, the monthly change in counts and the relative frequency of each interaction type were computed. These features–proportion of interactions that were conflictual versus cooperative, and how sharply events changed from the previous month–are used as predictors in the classification procedure.

[Figure 1 about here.]

MID hostilities are measured on a five-point scale, which was collapsed into a binary with the cutoff set at four. Events above this threshold involve the use of force, typically associated with "war," while events below this threshold require only the threat or display of force [9]. Further classification of exact hostility levels will be attempted in a later stage of this project.

2.3 Model

The mathematical model for this project is that a binary indicator of conflict, y, between country i and country j at time t is a function of observed interactions between them in month t. Formally,

$$\hat{y}_{i,j,t}|x \quad = \quad f(\Delta x_{i,j,t} + z_{i,j,t})$$

The conflict indicators $y_{i,j,t}$ are binary $(0, 1)$. The observations $x_{i,j,t}$ consist of the month-to-month change in interactions between i and j within each of the event categories described above ($\Delta x_{i,j,t} = x_{i,j,t} - x_{i,j,t-1}$). Thus, x is a count variable that can take on positive or negative values ($x \in \mathbb{Z}$). The observations $z_{i,j,t,k}$ measure the relative frequency of interactions in category m as a percentage of the total n observations for the dyad-month:

$$z_{i,j,t,m} = \frac{\sum_{k=1}^{n} x_{i,j,t,k} \mathbb{I}(x \in m)}{\sum_{k=1}^{n} x_{i,j,t,k}}.$$

Both the x and z values are derived from observations the GDELT data. The indicator of conflict, y, is observed in the MID data, and predicted indicators of conflict $\hat{y}$ will be estimated using the classification tree. The predicted values $\hat{y}$ for the test set can be compared to the actual MID data to validate the model. This will give us a sense of how accurate the classifications for post-2001 data will be. Even though these values will not be perfectly accurate, they should give us a good approximation of which countries experienced conflict since 2001 and can help speed up the production of the next generation of MID data.

2.4 Machine Learning Method

The inference problem is to compute a function $f(\cdot)$ that uses country interactions to estimate an indicator of whether conflict occurred. To accomplish this, this project will use a classification tree (CART). This method is appropriate for binary classification with real-valued predictors, which makes it well suited for this project.

CART grows a tree T from a root node by adding splits based on features. At each stage, there is an impurity function $I(T)$ that measures the (in)accuracy of the existing classifications. When adding another split to the tree, the CART algorithm seeks to maximize the impurity reduction ΔI. The measure of impurity reduction is supplied to the algorithm, as is a complexity parameter, κ. We can think of κ as the cost of adding another variable to the tree. When $\kappa = \infty$ the tree will consist of only the root node, and as we reduce κ toward zero we allow the tree to grow. However, setting κ too low can result in overfitting, which can be assessed via cross-validation. For more on classification trees in general and the implementation used for this project, see [17, 18, 19].

3 Results

3.1 Model Estimation

To fit the classification tree, this project used the `rpart` library in $\mathcal{R}$ [19]. The minimum complexity parameter, κ, was initialized to be 10^{-4}. By cross-validation, the optimal κ was found to be 0.00155 (see Figure 2). The Gini coefficient was used as the impurity function.

This value was used to prune the tree from its maximum of 176 splits down to nine. The resulting tree is shown in Figure 3, with leaves shaded by whether MID hostilities ("war") or peace is more likely. Each leaf also indicates the relative frequency of war in the training set for observations assigned to that leaf.

[Figure 2 about here.]

[Figure 3 about here.]

3.2 Interpretation

All of the twenty event types in Table 1 were supplied as candidate variables to the model, as were the percentage of observations within each quadrant of the table (i.e. material cooperation, verbal cooperation, material conflict, and verbal conflict). Of these, seven variables are included in the final tree. Investigations feature prominently in the first three splits. At first this seemed curious, given that investigations are classified as material cooperation by CAMEO (compare node eight). Upon closer examination, this category includes investigations of crime, corruption, human rights abuses, military action, and war crimes. Given that humanitarian interventions feature prominently in the MID dataset during the period under consideration, the influence of this variable becomes

3

less surprising. For example, NATO efforts in the Balkans involve a large number of dyads involved in the investigation of war crimes and other human rights violations.

The other splits are less surprising. Demands (node four) are likely to be more associated with domestic conflicts, such as protest movements, rather than interstate disputes. Large amounts of aid make conflict less likely (node seven), as would be expected. Unconventional violence (e.g. mass killings and ethnic cleansing) and the use of force (e.g. fighting with small arms, artillery, or aircraft) are associated with MIDs for obvious reasons. Overall the model seems to match the types of conflict observed in the 1990s, but might do less well in later periods.

3.3 Model Diagnostics

How does the tree perform relative to other classification models? Table 2 compares the tree above to a null model (all observations are predicted to be peaceful) and a logistic regression using the same features as the CART model. The classification tree outperforms the other two models in the training data (1992-1998, $n = 372, 271$). For the test data (1999-2001, $n = 213, 218$), the tree has the same mean-squared error (MSE) as the null model, but better precision and recall. Compared to logistic regression, the tree has worse recall but much better precision for the test data. This suggests that the classification tree generated more false negatives (i.e. missed the occurrence of some conflicts) in the test data.

[Table 2 about here.]

4 Conclusion

Both the accuracy and the interpretability of classification trees made them very useful for this project. In addition, the comparison to logistic regression shows that CART handled dependencies between the features better than GLM. The interpretation of the tree suggests that it fits the 1990s very well, but may perform less well in the future. Perhaps an ensemble of trees based on a few months of data at a time (with the addition of more trees as the data extend in to the future) would perform better and account for changes in the data generating process over time. A Bayesian ensemble of several classification trees could improve performance while still maintaining more interpretability than is available in random forests [10, 20, 21, 22].

Other adjustments can also help to refine the model at later stages. International conflict is a relatively rare event, meaning that in k-fold cross validation it is possible that some subsets will have no instances of conflict; to prevent this, synthetic minority over-sampling (SMOTE) could be used [23]. To incorporate interdependencies not captured at the dyadic level, future iterations could also include lags that measure conflict in social or spatial neighbors [24, 25, 26, 27, 28, 29].

The findings thus far indicate that automated production of political indicators is feasible to within a close approximation. The application of classification trees offers a quick and inexpensive tool that can save much human effort. A final pass by human coders will still be required to obtain fully accurate classifications, but the cost can be substantially reduced. With the additional steps proposed above, the process of automating political indicators using event data could also be applied to other widely used indices such as the Polity and Freedom House regime scales (measuring democracy and autocracy).

References

[1] M.D. Ward, B.D. Greenhill, and K.M. Bakke. The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research*, 47(4):363–375, 2010.

[2] Deborah J Gerner, Philip A Schrodt, Ronald A Francisco, and Judith L Weddle. The analysis of political events using machine coded data. *International Studies Quarterly*, 38(1):91–119, 1994.

[3] Deborah J. Gerner, Philip A. Schrodt, Yilmaz Ömür, and Rajaa Abu-Jabr. Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for a Post Cold War World. Boston, MA, August, 29-September 1 2002. Annual Meetings of the American Political Science Association.

[4] Gary King and Will Lowe. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, 57(03):617–642, 2003.

[5] Andrea Ruggeri, Theodora-Ismene Gizelis, and Han Dorussen. Events data as bismarck's sausages? intercoder reliability, coders' selection, and data quality. *International Interactions*, 37(3):340–361, 2011.

[6] P. Schrodt and K. Leetaru. Gdelt: Global data on events, location and tone, 1979-2012. *International Studies Association*, 2013.

[7] Sean P Obrien. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87–104, 2010.

[8] Philip A Schrodt. Tabari: Textual analysis by augmented replacement instructions, version 0.7. 2009.

[9] Faten Ghosn, Glenn Palmer, and Stuart A Bremer. The mid3 data set, 1993–2001: Procedures, coding rules, and description. *Conflict Management and Peace Science*, 21(2):133–154, 2004.

[10] Bryan Arva, John Beieler, Bejamin Fisher, Gustavo Lara, Philip A Schrodt, Wonjun Song, Marsha Sowell, and Sam Stehle. Improving forecasts of international events of interest. In *EPSA 2013 Annual General Conference Paper*, volume 78, 2013.

[11] James E. Yonamine. Working with event data: A guide to aggregation choices. *Ph.D. Thesis*, 2013.

[12] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1976.

[13] Beata Beigman Klebanov, Daniel Diermeier, and Eyal Beigman. Lexical cohesion analysis of political speech. *Political Analysis*, 16(4):447–463, 2008.

[14] Philip A Schrodt. Predicting interstate conflict outcomes using a bootstrapped id3 algorithm. *Political Analysis*, 2(1):31–56, 1990.

[15] N Speybroeck. Classification and regression trees. *International journal of public health*, 57(1):243–246, 2012.

[16] Robert Trappl, Johannes Fürnkranz, and Johann Petrak. Digging for peace: Using machine learning methods for assessing international conflict databases'. In *ECAI*, pages 453–457. PITMAN, 1996.

[17] Kevin P Murphy. *Machine learning: a probabilistic perspective*. The MIT Press, 2012.

[18] L Breiman JH Friedman RA Olshen and Charles J Stone. Classification and regression trees. *Wadsworth International Group*, 1984.

[19] Terry M Therneau and Elizabeth J Atkinson. An introduction to recursive partitioning using the rpart routines. Technical report.

[20] Jacob M Montgomery, Florian M Hollenbach, and Michael D Ward. Improving predictions using ensemble bayesian model averaging. *Political Analysis*, 20(3):271–291, 2012.

[21] Adrian E. Raftery. Bayesian model selection in social research (with discussion). In Peter V. Marsden, editor, *Sociological Methodology 1995*. Blackwell, Cambridge, MA, 1995.

[22] Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. Using bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5):1155–1174, 2005.

[23] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357, 2002.

[24] Kristian S Gleditsch and Michael D Ward. War and peace in space and time: The role of democratization. *International Studies Quarterly*, 44(1):1–29, 2000.

[25] Kristian S Gleditsch and Michael D Ward. Measuring space: A minimum-distance database and applications to international studies. *Journal of Peace Research*, 38(6):739–758, 2001.

[26] Peter D Hoff and Michael D Ward. Modeling dependencies in international relations networks. *Political Analysis*, 12(2):160–175, 2004.

[27] Michael D Ward and Kristian S Gleditsch. Democratizing for peace. *American Political Science Review*, pages 51–61, 1998.

[28] M.D. Ward, R.M. Siverson, and X. Cao. Disputes, democracies, and dependencies: A reexamination of the kantian peace. *American Journal of Political Science*, 51(3):583–601, 2007.

[29] Michael D Ward, Katherine Stovel, and Audrey Sacks. Network analysis and political science. *Annual Review of Political Science*, 14:245–264, 2011.
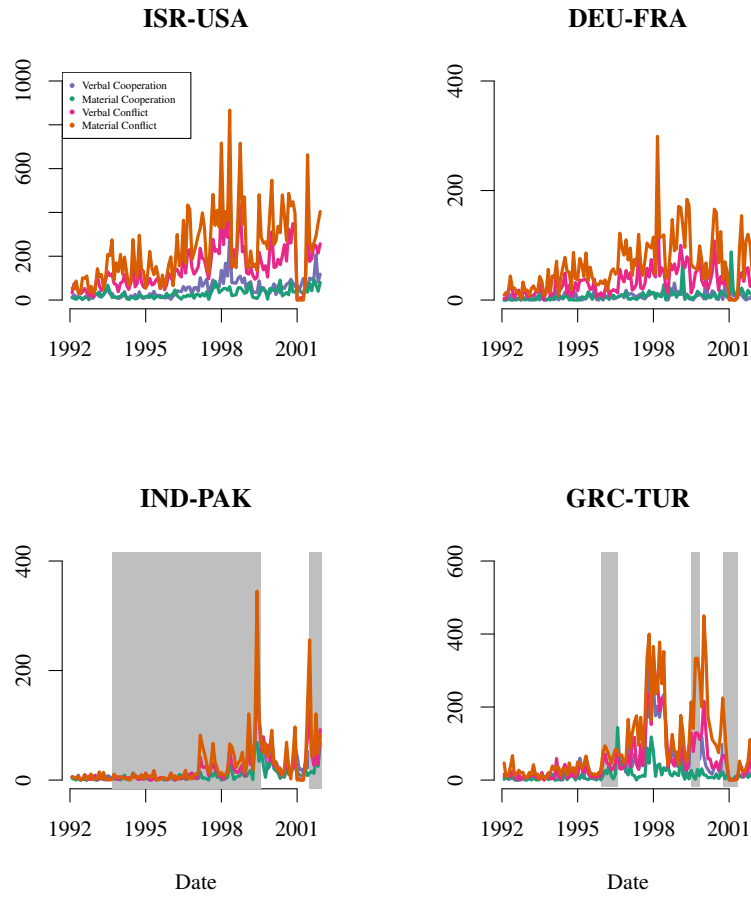
Figure 1: Examples of Dyadic Time Series Derived from GDELT, with and without MIDs (shaded)
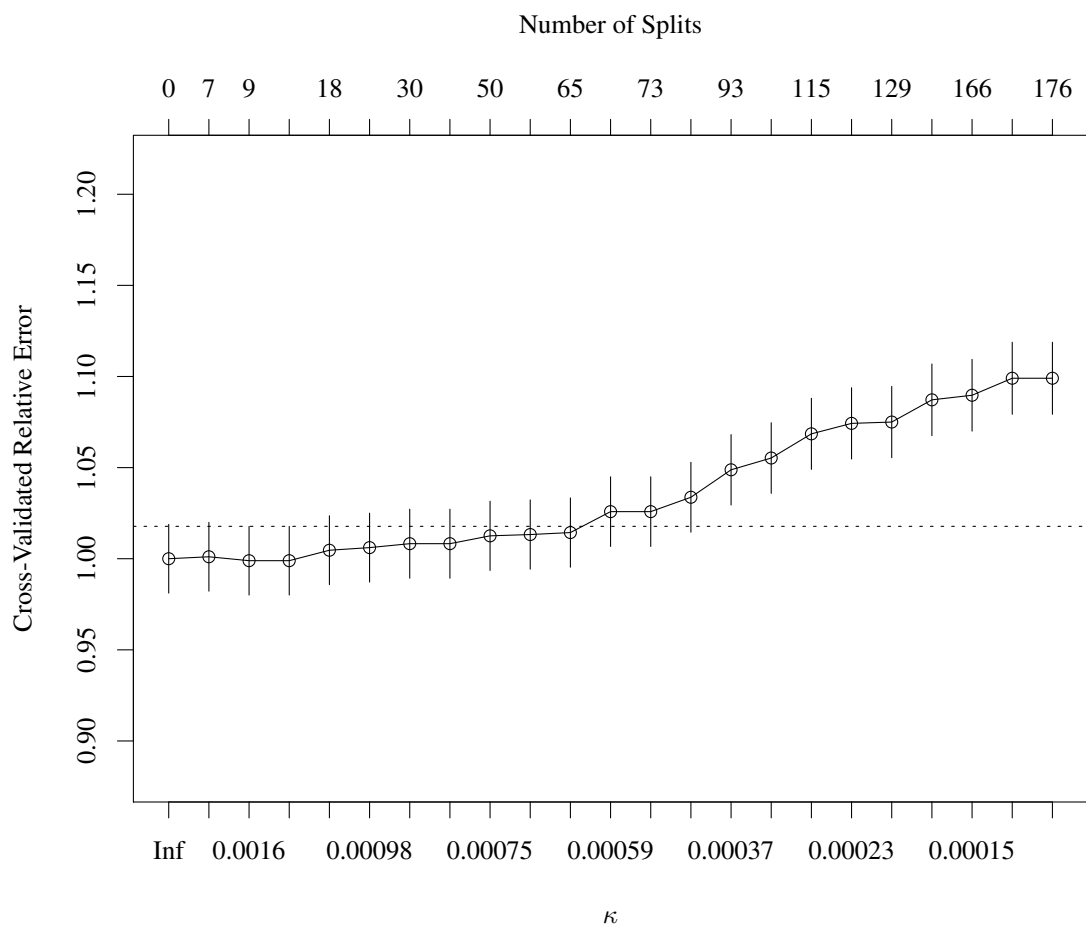
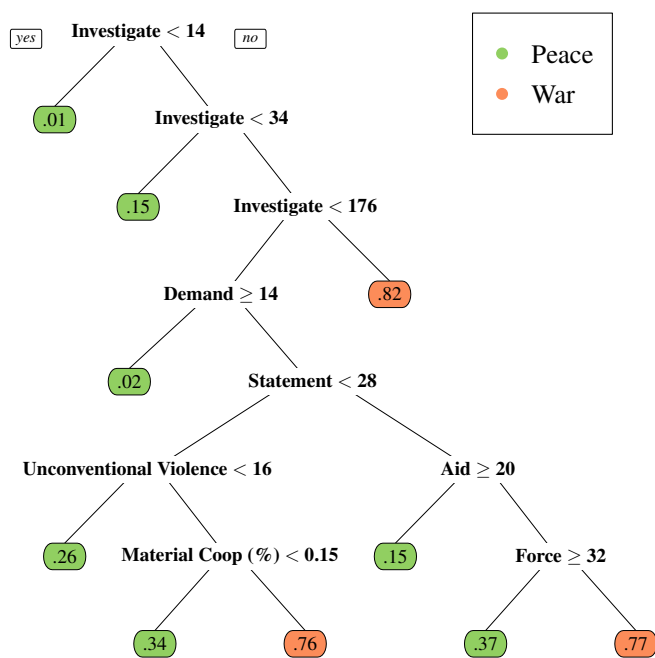Figure 2: Cross-Validation Error and Number of Splits as a Function of κ

Figure 3: Fitted and Pruned Classification Tree

Table 1: CAMEO event categories and descriptions

	Cooperative	**Conflictual**
Verbal	public statement, appeal, express intent to cooperate, consult, engage in diplomatic cooperation	demand, disapprove, reject, threaten, protest
Material	cooperate materially, provide aid, yield, investigate	exhibit force posture, reduce relations, coerce, assault, fight, use conventional mass violence

Table 2: Comparison of CART to Alternative Models

Model	Training Data			Test Data		
	MSE	Precision	Recall	MSE	Precision	Recall
Null	0.0075	0.000	0.000	0.0066	0.000	0.000
GLM	0.0082	0.158	0.022	0.0079	0.142	0.038
CART	0.0067	0.702	0.192	0.0066	0.422	0.027