Predicting Intrastate Conflict with AI

Ryan Holmdahl and Stephanie Chen

October 21, 2015

1 Purpose

Violent conflict, being expensive in both life and treasure, should be avoided if at all possible. For preventative measures to be enacted, policymaking institutions must be able to predict where and when violence is likely to emerge. While the development of *interstate* conflict is complicated by the intricacies of policy, treaties, and international law, we believe predicting the onset of *intrastate* conflict can be reasonably accomplished by examining existing data on individual nations through the lens of artificial intelligence. We endeavor to devise a solution by which one can determine whether or not such a conflict is likely to emerge in a given state in a given year.

2 Input-Output Examples

Suppose we want to know if civil war is likely to break out in Finland in 2016. We can write this request as (Finland, 2016), signifying that we want to predict conflict in the year 2016 in Finland. Our algorithm would, using economic, social, historical, and political data about Finland, return a value indicative of whether such a war is likely. In this case, the algorithm would likely output -1, as an intrastate conflict in Finland is improbable at this time.

Similarly, if we wanted to predict whether or not violent conflict will occur in Mexico in 2017, we could input (Mexico, 2017). In this case, we might receive a 1 as output if the algorithm deemed such a conflict probable.

We would also be able to make predictions on years that have already happened, primarily for testing purposes. For example, if we input (Serbia, 1992), we would hope to see a 1 returned, as an intrastate conflict did occur in Serbia in 1992.

3 Baseline

For our baseline, we use a simple majority-label algorithm on the union of two datasets: the UCDP Non-State Conflict Dataset [1], which records conflicts in which a state was not represented, and the UCDP/PRIO Armed Conflict Dataset [2], which records conflicts in which at least one state was involved. For every state represented in the data and every year for which there is data (1989-2014), we predict that an intrastate conflict would not emerge. This resulted in a prediction accuracy of roughly 77.2%, a high baseline but one we believe we can improve upon.

4 Oracle

Our oracle is simply a consultation of the conflict databases. For any given country in any given year, we can simply check to see if a conflict did break out. For any historical case, our oracle has a 100% chance of returning the correct answer, assuming as we do that the databases are valid. The disparity between our oracle and our baseline is fairly small, as one might suspect, given that intrastate conflict is not an especially frequent occurrence.

For future cases, there is no actual means of peeking at the correct answer, but reviewing scholarly literature on a country and its future might yield some interesting comparisons with our own findings.

5 Existing Literature

Much work already exists in political science and related fields regarding the factors involved in the outbreak of violent intrastate conflict. Especially in poorer and developing countries, the presence of class inequalities in political, economic, and social power or access is a strong indicator of intrastate conflict, with underserved groups often turning to violent action[3][4]. Slowed economic growth or economic stagnation, high unemployment, and a lack of government services are also contributing factors[3]. Contrary to what might be expected, ethnic or religious diversity is not strongly indicative of civil war, with factors like population size and terrain being more significant[6].

Despite the abundance of literature on conflict factors, as well as the presence of quantitative measures of most factors, the applications of artificial intelligence to prediction of conflict—especially intrastate conflict—are underexplored. The Artificial Intelligence for the Avoidance of Crises and Wars (AISOC) project[7] has done perhaps the most in this area, conducting analysis on conflict and mediation databases using AI techniques. Their work includes decision tree learning and generation from the CONFMAN conflict management database[8], nearest-neighbors classification of various conflict-mediation cases in the KOSIMO political conflict database[9], and algorithmic pattern searching for time dependencies in conflicts[10]. Some work has also been done in predicting interstate conflict using neural networks and support vector machines (SVMs) in a similar (though more complex) approach to ours[11]. In the existing literature, however, the specific problem of modeling and predicting the development of intrastate conflict has not yet been addressed in research.

6 Approach

We plan to approach this issue from a machine learning perspective. The program accepts as input a country and a year. If an intrastate conflict is expected to occur in the given year in the given country, 1 is returned; otherwise, -1 is returned. Naturally, the farther the input year is from the present, the less accurate this algorithm would be, as we would have to extrapolate our known data further.

The feature vector for each input will model those indicators found by existing research to correlate to conflict, potentially including but certainly not limited to income inequality, population size, poverty rate, economic stagnation, and unemployment. We will examine data for the given year and the years before in an attempt to establish future trends as well as assess the nation's status at the given time. Public United Nations databases contain most, if not all, of the information we will need to construct these features, though some may be harder to quantify than others (see following section).

To create our weights vector, we will train on a dataset compiled from intrastate conflict databases, including the UCDP Conflict Encyclopedia and The Correlates of War, to determine the impact each feature has on the breakout of violent conflict. We can similarly extract cases from those databases to serve as our testing set to determine if our algorithm accurately predicts the emergence of conflict.

7 Challenges

To select features, we can consult the substantial body of research on the factors that precipitate conflict. Even so, we will want to explore beyond these statistics to determine if certain value thresholds are relevant or if some features have disproportionate importance.

We are also forced to address inconsistent information. For example, Afghanistan has only one year of unemployment data available, while Albania has over 20. We must devise our algorithm such that this data can be used when present without rendering analysis of more limited datasets excessively inaccurate. In our previous example, if we were to include a feature like "unemployment five years ago," the algorithm would become worse at classifying Afghanistan. We can hopefully circumvent this with quantitative interpretations of the data rather than the raw numbers. A "declining unemployment" feature, with a value of 1 if declining, -1 if increasing, or 0 if unavailable, might be more effective in maintaining accuracy on both countries, as it allows for a neutral score in the absence of data over time.

Another issue we are likely to encounter are the gaps between the extent of our data and the date of prediction. Consider a country whose data ends in 2006. If we wanted to predict the likelihood of conflict in 2009, we would somehow have to adjust for the missing years of data. A weight vector built on full data would not work properly in this case, as it would be learned on a full dataset and therefore weighted with the assumption that there are no gaps. The most straightforward solution would be to interpolate the gap using the trends in the available data. Alternatively, we can create different weight vectors which are trained on missing information. A weight vector for a gap of three years, for example, would in training not take into consideration the three years of data before each analyzed year.

In addition, because so many of our indicating factors are somewhat subjective in and of themselves, the numerical values we use as features will be inherently incomplete—for example, a simple index value like the Gini coefficient cannot sufficiently describe the degree and scope of a country's income inequality. We will need to conduct an in-depth analysis of measures used in prior work to determine if they are reasonable indicators or if they would otherwise hurt the credibility of our project.

References

[1] Sundberg, Ralph, Kristine Eck and Joakim Kreutz, 2012, "Introducing the UCDP NonState Conflict Dataset", *Journal of Peace Research*, March 2012, 49:351-362. Version 2.5 - 2015.

[2] Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research* 39(5).

Latest presentation by Pettersson, Therése & Peter Wallensteen (2015) Armed Conflicts, 1946-2014. *Journal of Peace Research* 52(4).

[3] Stewart, F. (2002). Root causes of violent conflict in developing countries. BMJ: British Medical Journal, 324(7333), 342–345.

[4] Caprioli, M. (2005), Primed for Violence: The Role of Gender Inequality in Predicting Internal Conflict. International Studies Quarterly, 49: 161–178. doi: 10.1111/j.0020-8833.2005.00340.x

[5] Harff, B., & Gurr, T. Robert. (2004). Ethnic conflict in world politics. 2nd ed. Boulder, Colo.: Westview Press.

[6] ”Ethnicity, Insurgency, and Civil War”,”JAMES D. FEARON”,”DAVID D. LAITIN”,American Political Science Review,Volume 97,Issue 01,2003,pp 75-90,DOI 10.1017/S0003055403000534

[7] Trappl, R. (PI) ”Artificial Intelligence for the Avoidance of Crises and Wars.” http://www.ofai.at/research/aisoc/peace/

[8] Trappl, R. and Fürnkranz, J. and Petrak, J., Digging for Peace: Using Machine Learning Methods for Assessing International Conflict Databases, In Proceedings of the 12th European Conference on Artificial Intelligence (ECAI-96, 1996, 453–457, John Wiley.

[9] Fürnkranz, J. and Petrak, J. and Trappl, R., Knowledge Discovery in International Conflict Databases, Applied Artificial Intelligence, 1997, 11, 91–118.

[10] Kovar K., Fürnkranz J., Petrak J., Pfahringer B., Trappl R., Widmer G., 2000. Searching for Patterns in Political Event Sequences: Experiments with the KEDS Database, Cybernetics and Systems, 31(6)649-668.

[11] Habtemariam, E. and Lagazio, M. Artificial Intelligence for Conflict Management, unpublished, http://arxiv.org/pdf/0705.1209.pdf.