**Can Machines Learn to Predict a Violent Conflict?**

**Chris Perry**

The need for comprehensive early warning systems to analyze and disseminate data on sociopolitical and armed conflict dynamics within the UN system is well established. For at least the last two decades, there have been calls within the United Nations to develop robust, accurate, and effective early warning system for conflict prevention. Indeed, as recently as September 2011, Secretary-General Ban Ki-moon and the UN Security Council reiterated this need once again through their [report](http://www.un.org/wcm/webdav/site/undpa/shared/undpa/pdf/SG%20Report%20on%20Preventive%20Diplomacy.pdf) “Preventative Diplomacy: Delivering Results.” The president of the Security Council at the time [stated](http://www.un.org/News/Press/docs/2011/sc10392.doc.htm) that a “key component…of a comprehensive conflict prevention strategy include[s] early warning [mechanisms].”

One of the major operational challenges to early warning is simply aggregating incoming information and data to derive actionable intelligence on an emerging situation. Often (but not always), incoming data is highly qualitative, which can place strains on the limited capacity of international organizations (IOs) and non-governmental organizations (NGOs), which is often limited. In addition, quantitative data is often not collected in a way that can easily be fed into a larger system. Organizations can find it too resource-intensive to clean, process and analyze the data, thus limiting the type and volume of data being looked at.

One way to overcome these resource constraints is to create tools that can automate the processing and analysis of quantitative data. Machine learning and data science seems a natural fit to improve this process. Data science is a multidisciplinary field that applies a mix of mathematics, statistics, computer science, data modeling and visualization, graphic design and hacking, as well as specific subject area expertise. Machine learning is a branch of computer science that leverages algorithms, or a set of step-by-step computer procedures, to perform actions without explicitly being programmed to. Machine learning has been used by a wide variety of private sector organizations for things like targeting user recommendations, detecting fraud and identity theft, and ad optimization.

Automated early warning systems can help NGOs and IOs in a number of ways. They can help organizations develop an evidence base to create the political will to do preventative work to intervene or mitigate negative effects of large-scale conflict as tensions ramp up. In the case of predicting conflict, organizations can use early warning risk assessments for better planning and try to target non-conflict interventions that have conflict-mitigating knock-on effects in high-risk areas.

Yet there are relatively few examples of systematic attempts to create open source tools to forecast violent conflict. Instead, existing efforts to use statistical forecasting are 1) classified, 2) proprietary and very expensive, or 3) rudimentary, often relying heavily on data of violent occurrences as the primary source of information about trends in violence. Classified data can only be used by NGOs and IOs on a need-to-know basis. The costs of proprietary systems are often prohibitive to IOs and especially to NGOs.

That leaves rudimentary systems as the only viable option for many organizations. Yet rudimentary efforts don’t need to remain so. There is a huge array of open source tools that can be used to build the necessary components to a vastly improve rudimentary systems.

As part of IPI’s new Data Lab project, we have been looking at ways to leverage data science methods into our policy research on peace, security, and conflict prevention. One area of research over the last year has been to research the application of machine learning specifically to the conflict prevention and early warning problems. (The [results](http://www.stabilityjournal.org/article/view/sta.cr) of the first stage of the project were just published in the *Stability Journal’s* special collection on new technologies for peace and development.)

The first stage focused on two main aspects: feasibility and added value. First, what are the necessary systems for a feasible technology-based early warning tool? Can these systems be automated so as to make the tool workable, and, if so, what is the process? Is there a way to integrate a variety of types of data into the system?

Second, does the use of machine learning actually add any predictive capabilities above and beyond a baseline of knowledge that an area had experienced violence in the recent past?

Our initial results were promising. We developed a framework that aggregates a variety of sources at the subnational level and tries to predict outbreaks in violence. This sub-national unit of observation is an important step. It is relatively easy to say with a reasonable amount of certitude that the Democratic Republic of the Congo or Somalia will experience conflict in the coming year based on knowledge that those countries have experienced high levels of violence in the recent past. It is much harder, and potentially much more useful, to be able to pinpoint the specific areas in a country that are likely to experience conflict.

We also found that the application of a machine-learning algorithm improved our predictive accuracy. One of the problems with trying to predict outbreaks of violence is that violent events are very rare relative to the larger data set. Because of that, there is very little variation in overall accuracy, either due to algorithm choice or due to use of additional data.

Using only baseline data (knowing that a district had previously experienced conflict in previous years), our algorithm produced a very high overall accuracy rate. Almost all of the errors were in over-predicting conflict. In other words, all of these models tend to over-predict towards false positives rather than false negatives. Because of this, the accuracy rate related to these false positives is the one to watch for performance gains.

Indeed, using the full data1 over the baseline of only previous conflict offered a modest but promising 10-30% increase in accuracy. While this may seem less than groundbreaking, the input data was selected primarily for ease of use in the test model. Machine learning is an iterative process, and future iterations will focus on selecting better input data.

As a final “sanity check”, the models were applied to 2012 data, which was initially partitioned from the rest of the dataset. The predictions were then mapped against actual instances to give a visual representation of the accuracy of the models. These maps give a good visual representation of the geographically accuracy of both learners. Probably the most promising result is that while the models tend to over-predict; they do not do so randomly. Over-prediction values tend to be geographically clustered in the “right” place.

These initial results show promise and offer one potential path towards an open source early warning tool for IOs and NGOs. That being said, there is a long way to go from the toy model to a functional tool.

1The full predictor data we used included raster data (a land conflict index, flood and drought frequency, percentage of children under five years of age that are underweight, infant mortality rate, GDP of the current and previous two years, change in GDP, and the population for the current year and the previous two years), GIS vector data (ethnic composition and petroleum resources), GIS point-data (lootable diamond deposits and conflict events), and national and subnational governance data.