Computer forecasting matched rises, ebbs in Afghanistan violence

July 17 2012, By Jon Bardin

In August 2010, shortly after WikiLeaks released tens of thousands of classified documents that cataloged the harsh realities of the war in Afghanistan, a group of friends - all computer experts - gathered at the New York City headquarters of the Internet company Bitly Inc. to try and make sense of the data.

The programmers used simple code to extract dates and locations from about 77,000 incident reports that detailed everything from simple stop-and-search operations to full-fledged battles. The resulting map revealed the outlines of the country's ongoing violence: hot spots near the Pakistani border but not near the Iranian border, and extensive bloodshed along the country's main highway. They did it all in just one night.

Now one member of that group has teamed up with mathematicians and computer scientists and taken the project one major step further: They have used the WikiLeaks data to predict the future.

Based solely on written reports of violence from 2004 to 2009, the researchers built a model that was able to foresee which provinces would experience more violence in 2010 and which would have less. They could also anticipate how much the level of violence went up or down.

The project, whose results were published online Monday by the Proceedings of the National Academy of Sciences, is part of a growing movement to understand and predict episodes of political and military...
conflict using automated computational techniques.

The availability of huge amounts of data combined with steady increases in computing power has prompted experts to bring the rigor of objective quantitative analysis to realms that were once considered fundamentally subjective, including literature and the study of social groups.

"For the first time, we have large data sets from places like Facebook and Twitter that we can analyze with high-powered computers and get meaningful results," said Paulo Shakarian, a computer scientist at the U.S. Military Academy at West Point, who is working on an algorithm to predict the location of insurgent weapons caches. "Iraq and Afghanistan are the very first conflicts where we have been collecting as much data as we possibly can."

In the case of the WikiLeaks data, the researchers sought to find a general pattern to the violence in Afghanistan and use it to predict how it would change in each province in 2010 - the year President Barack Obama increased the number of U.S. troops in the country.

"The model we employed is both complex and simple," said Guido Sanguinetti, an expert in computational sciences at the University of Edinburgh in Scotland and the study's senior author. "It doesn't take in any knowledge of military operations or political events, and it treats all types of violence exactly the same, whether it's a stop-and-search or a big battle."

Even with these ostensibly key details missing, the researchers found that they could predict 2010's events with striking accuracy.

And the model wasn't tripped up by Obama's decision to send 30,000 additional troops, which introduced a new dimension to the Afghanistan
conflict.

"Our findings seem to prove that the insurgency is self-sustaining," Sanguinetti said. "You may throw a large military offensive, but this doesn't seem to disturb the system."

The study authors said they were most surprised that the model could predict activity even in Afghanistan's relatively quiet northern provinces, where there were fewer data points available to analyze.

"This shows that the escalation we see isn't just attributed to the noise in the data," said study leader Andrew Zammit Mangion, a computational sciences researcher at the University of Edinburgh. Instead, he said, patterns existed nearly everywhere.

Michael Ward, a political scientist at Duke University who has shown that location data can improve predictions of conflicts, said the study pointed the way to future research.

"Suppose you could say, 'This is the effect on violence if you build different types of infrastructure,' " he said. "They don't do that, but they've set up the framework to do it."

The study also shows why it's important to make as much data public as possible, Ward said. Without WikiLeaks, he said, a study like this would have been far more difficult to carry out.

Clionadh Raleigh of Trinity College Dublin, who uses data to predict violence in Africa based on factors such as the outcomes of local elections, said the Afghanistan model could be made even better by including variables such as the political party in power.

"Violence, in general, is a really good predictor of future violence," she
said. But even better would be "to figure out what stops the cycle of conflict."

Quantitative rigor is making its way into some surprising fields of study. In 2010, a few months after WikiLeaks' data dump, Google released a database of every single word contained in thousands of books published between 1800 and 2000 - about 4 percent of all books ever printed. That has enabled some intrepid researchers to close in on the final frontier: Studying literature with advanced math.

In a study published last year in Science, experts from Harvard University and Google were able to detect evidence of censorship regarding controversial historical figures and events, such as early Soviet official and Stalin foe Leon Trotsky and the 1989 Tiananmen Square massacre in China.

"That's what this digital humanities focus is being driven toward: uncovering trends in data that have just never been available before," Raleigh said.

More information: Point process modelling of the Afghan War Diary, PNAS, Published online before print July 16, 2012, doi: 10.1073/pnas.1203177109, www.pnas.org/content/early/201...1203177109.abstract

(c)2012 Los Angeles Times
Distributed by MCT Information Services
