Interpretable machine learning predicts terrorism worldwide

August 2 2021
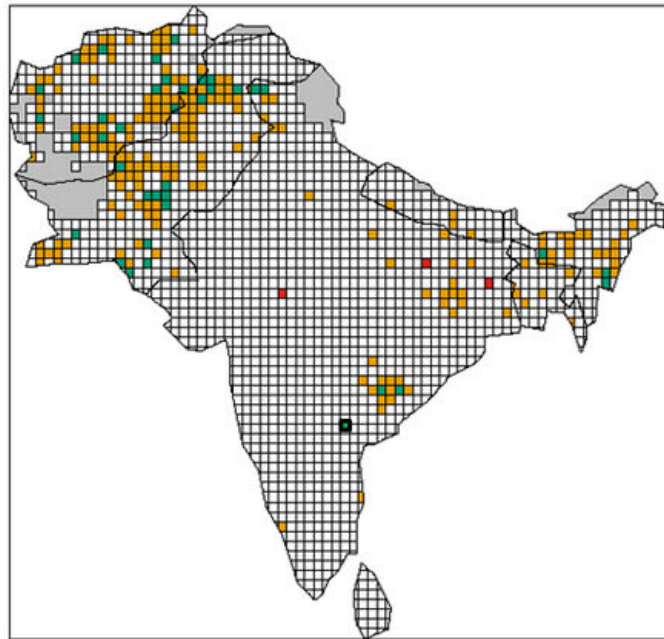

IMAGE: SNAPSHOT OF A WEEKLY PREDICTION OF NON-STATE TERRORISM IN GRID CELLS (50 KM X 50 KM) ACROSS SOUTH ASIA

(■TRUE POSITIVE; □ TRUE NEGATIVE; ■FALSE POSITIVE; ■FALSE NEGATIVE).

CREDIT: ANDRE PYTHON (2021)

Credit: Zhejiang University

About 20 years ago, a series of coordinated terrorist attacks killed almost 3,000 people in the World Trade Center, New York and at the Pentagon. Since then, a vast amount of research has been carried out to better

understand the mechanisms behind terrorism in the hope of preventing future potentially devastating acts of terror. Despite the large efforts invested to study terrorism, quantitative research has mainly developed and applied approaches aiming at describing regional cases of terrorist acts without providing reliable and accurate short-term predictions at local level required by policymakers to implement targeted interventions.

Building a model to predict terrorism worldwide at fine spatiotemporal scales

Publishing in *Science Advances*, an international research team led by Dr. Andre Python from the Center of Data Science at Zhejiang University investigate [machine learning algorithms](#) capable of predicting and explaining at fine spatiotemporal scale the occurrence of [terrorism](#) perpetrated by non-state actors outside legitimate warfare (non-state terrorism) across the world. To cover all regions worldwide potentially affected by terrorism over a large time period, the authors consider about 21 million week cells, which are composed of 26,551 grid cells at 50 km × 50 km that cover inhabited areas in the world over a period of 795 weeks between 2002 and 2016. An interpretable tree-based machine learning algorithm is compared with alternative benchmark [predictive models](#) to predict and explain the probability of the occurrence of terrorism (response) in each week cell across the world. Informed by terrorism theory, the model includes 20 structural features—time-invariant variables that account for the effect of, e.g., per capita gross domestic product (GDP)—and 14 procedural features—dynamic variables that account for the fact that terrorism activity in the past affects the risk of terrorism in the future. To predict complex social phenomena such as terrorism at fine spatiotemporal scales, theoretically informed machine learning algorithms are likely to outperform parsimonious models using procedural features only, says Dr. Andre

Python who led the research. The choice of the features included in the predictive model is crucial; the relevance of the model outputs and the predictive performance benefit from a solid conceptual understanding of the mechanisms driving terrorism at the scale on which predictions are made.

Can terrorism be accurately predicted?

While the predictive performance of machine learning algorithms is relatively high in areas that are highly affected by terrorism, it remains challenging to predict events that occur in regions that have not experienced terrorism over a long period. Algorithms may show a relatively good overall accuracy even at fine spatial and temporal resolution. However, it is virtually impossible to predict 'black swan events'—those events that occur only once over a very large period of time, says Python. Terrorist events occurred in less than 2% of the week cells considered in our global study. Data imbalance reduces the precision of the models, which is the number of week cells that encountered terrorism and have been correctly predicted divided by the total number of week cells predicted to encounter terrorism. This means that to prevent a large proportion of terrorist events in a region that is not much affected by terrorism, important resources are required to survey large areas where terrorism can potentially occur.

Along with disagreement among scholars about the definition of terrorism, the availability, spatiotemporal coverage, and the quality of publicly available data on terrorism and its potential drivers remain an important barrier to accurately predict terrorism globally and at policy-relevant scales, says Python. But terrorism data and socioeconomical drivers are becoming more detailed, comprehensive and more easily accessible. Also, the ongoing development in interpretable machine learning algorithms is very promising and will make these powerful tools more accessible to the research community and practitioners in the

coming years.

The important role of interpreting the results of machine learning algorithms

Until recently, the interpretation of models was almost essentially reserved to classical statistical models which impose a parametric relationship between features and the response like in linear regression models where features are assumed to be linearly associated with the response, and the coefficient associated with each feature can be estimated and further interpreted in line with existing terrorism theories. In this study, the researchers used an interpretable machine-learning algorithm to obtain relatively high predictive performance without compromising the interpretability of the results.

The research team used a gradient-boosted trees algorithm, from which they compute the accumulated local effect (ALE) plots, which highlight the marginal difference in the predicted probability of terrorism occurrence with an incremental change in the feature. The relationship between features and the occurrence of terrorism is likely to be non-linear and cannot be identified by standard statistical models, said Python. The ALE plots are an important interpretative tool. They can capture these complex relationships learned by the algorithm, says Python. In our study, we assessed the relationship between 34 relevant features and the occurrence of terrorism in 13 regions worldwide, he adds. We observed that some feature relationships are stable while others are more variable across regions. These results allowed us to better understand regional similarities and differences in the effects of major drivers of terrorism.

The machine-learning algorithm has potentially captured complex relationships of local and global drivers of terrorism at a scale that is

relevant for policy makers says Python. The interpretability of our model has important benefits beyond its predictive capabilities. Results can be analyzed in line with terrorism theories and can therefore contribute to build trust among modelers and practitioners, which is a crucial step to make these algorithms valuable for the entire research community.

More information: Andre Python et al, Predicting non-state terrorism worldwide, *Science Advances* (2021). DOI: 10.1126/sciadv.abg4778

Provided by Zhejiang University