



Applying Tensor Completion to the Study of Inter-State Competition and Irregular Conflict

Gray-zone conflict: broadening the aperture of what national security officials must deal with

Government agencies – in this case the US government specifically – need ways to interpret things out of open-source information that are not being directly reported. This becomes useful when talking about things such as propaganda and information campaigns. These “gray-zone” activities, so termed because they straddle the watershed between political competition and conventional warfare, are difficult to detect and discern from the usual noise of legitimate socio-political discourse. Previously, it was enough to know where the tanks, planes, and soldiers were. This was a scoped problem, which could be handled very well. With the emergence of gray-zone conflict, the aperture of what national security professionals must deal with has broadened considerably. Other emerging domains that make gray-zone activities more relevant and increasingly complex include the cyber domain and economic competition. Essentially: it’s easy to count a tank. It’s much harder to determine if something is fake news. Furthermore, threats develop more quickly and, due to modern technology, everything is networked; a single tweet can dethrone a politician and protests can start overnight. Responding to threats must be more agile and the information must be ready to present to decision makers in a way they can digest. This necessitates a unique approach to detecting gray-zone activities from open-source data. Furthermore, attempts to leverage non-data driven approaches for detecting gray-zone activities are influenced by human bias.

Reading between the lines: enabling decision makers to mitigate influences

VANE – the Virtual Anticipatory Network – (developed by KNIME Partner BigBear.ai) detects gray-zone activity from open-source data by reading in between the lines of directly reported phenomena. This allows decision makers to understand how US activities influence complex systems to mitigate gray-zone influences. It’s not designed to predict the precise date that an anticipated event is going to take place. VANE monitors data streams of known drivers that are predictive of the event of interest – and models courses of action to reduce the likelihood of the event’s occurrence. VANE helps give decision makers quantitative insights to do this. It’s a data-driven platform answering questions such as what does the future hold for X? Publicly available data such as demographic data, econometric data, news, social media, web information, and more is plugged into the system and provides the insights that decision makers need to achieve the desired end-state.

Company

The mission of the Combating Terrorism Technical Support Office is to identify requirements to combat terrorism and provide solutions to warfighters, first responders, and other front-line users as rapidly as possible. It’s vision is to Identify and develop capabilities to combat terrorism and irregular adversaries and to deliver these capabilities to DoD components and interagency partners through rapid research and development, advanced studies and technical innovation, and provision of support to US military operations.

- Tensor Completion
- US Army G2
- Government



"Most projects that the CTTSO works on require out of the box solutions, which means creative thinking is needed and in many cases, the need to extend the software - which is possible with KNIME. As a result, over 15 custom nodes have been written. This makes the projects easier to manage, because they are still being completed within a single software environment."

Curtis Fox, Advanced Analytics Team, Combating Terrorism Technical Support Office

Working with unclean data and determining the difference between fake news and real news

VANE currently leverages 14 open-source databases, tracking 660 independent metrics. The model spans many different topic areas – from economies to weather to cyber – of which many are not pristine. For example, one of the event sources is GDELT, which tracks 17,000 different news sources in 70 different languages. However, it struggles to maintain a clean dataset. Every time an Air Jordan sneaker is sold, an event in Jordan (the country) occurs. GDELT also doesn't determine between fake news, bad reports, and so on. Usually to compensate for low-quality data, data scientists must impute metrics or clean data-streams manually, which is time intensive and can introduce errors in the model.

Therefore, tensor completion is used. Specifically matrix factorization, which is a two-degree form of tensor completion. In this case, this has been extended into multiple domains because the real world isn't just users and ratings. It's countries and sensors and different time dimensions, meaning there is a lot more going into the tensor. The power of tensor completion is it finds relationships not just between two entities, but between the features that are being used to model the relationship between entities..

An example of how this works for image recognition: feed in some images that have pixel error, obfuscation, or where entire chunks of the image are missing entirely. The tensor completion algorithm will learn what it can about how pixels, edges, and colors relate within the image and provide a reconstructed view of what it thinks the data should look like. The other benefit is that a matrix is provided, which states where the suspected error is inside the data.

Project Results

VANE is currently undergoing BETA testing, and a production release of the software will be complete in 2020. Testing and evaluation has shown that VANE is effective at detecting information operations, cyber campaigns, and other means of indirect influence. VANE will be deployed to understand how gray-zone activities are being used to indirectly shape conflict zones in the 21st century.

Why KNIME?

KNIME Analytics Platform is great for many reasons. One of the greatest things about KNIME is that it's a no-coding workflow building environment. However, it's possible to add code when and where needed, which is what was done in this case. There are few developers and even less technical specialists working in the government. Therefore, giving them a platform that lowers the barrier to entry so that they can engage productively and communicate with the actual developers is important.

Most projects that the CTTSO works on require out of the box solutions, which means creative thinking is needed and in many cases, the need to

extend the software - which is possible with KNIME. As a result, over 15 custom nodes have been written. This makes the projects easier to manage, because they are still being completed within a single software environment. "We've got a lot of great things that help us take the value that KNIME gives us and fit it right in the little square peg, round hole that we're working in. And we can do it very, very quickly, plus integrate with all the other platforms that the government likes to use" says Brian Frutchey, VANE Technical Lead at BigBear.ai.