

Method finds hidden warning signals in measurements collected over time

December 17 2020



MIT researchers have developed a deep learning-based algorithm to detect anomalies in time series data. Credit: MIT News

When you're responsible for a multimillion-dollar satellite hurtling through space at thousands of miles per hour, you want to be sure it's

running smoothly. And time series can help.

A time series is simply a record of a measurement taken repeatedly over time. It can keep track of a system's long-term trends and short-term blips. Examples include the infamous COVID-19 curve of new daily cases and the Keeling curve that has tracked atmospheric carbon dioxide concentrations since 1958. In the age of big data, "time series are collected all over the place, from satellites to turbines," says Kalyan Veeramachaneni. "All that machinery has sensors that collect these time series about how they're functioning."

But analyzing those time series, and flagging anomalous data points in them, can be tricky. Data can be noisy. If a [satellite operator](#) sees a string of high temperature readings, how do they know whether it's a harmless fluctuation or a sign that the satellite is about to overheat?

That's a problem Veeramachaneni, who leads the Data-to-AI group in MIT's Laboratory for Information and Decision Systems, hopes to solve. The group has developed a new, [deep-learning](#)-based method of flagging anomalies in time series data. Their approach, called TadGAN, outperformed competing methods and could help operators detect and respond to major changes in a range of high-value systems, from a satellite flying through space to a computer server farm buzzing in a basement.

The research will be presented at this month's IEEE BigData conference. The paper's authors include Data-to-AI group members Veeramachaneni, postdoc Dongyu Liu, visiting research student Alexander Geiger, and master's student Sarah Alnegheimish, as well as Alfredo Cuesta-Infante of Spain's Rey Juan Carlos University.

High stakes

For a system as complex as a satellite, time series analysis must be automated. The satellite company SES, which is collaborating with Veeramachaneni, receives a flood of time series from its communications satellites—about 30,000 unique parameters per spacecraft. Human operators in SES' control room can only keep track of a fraction of those time series as they blink past on the screen. For the rest, they rely on an alarm system to flag out-of-range values. "So they said to us, "Can you do better?" says Veeramachaneni. The company wanted his team to use deep learning to analyze all those time series and flag any unusual behavior.

The stakes of this request are high: If the deep learning algorithm fails to detect an anomaly, the team could miss an opportunity to fix things. But if it rings the alarm every time there's a noisy data point, human reviewers will waste their time constantly checking up on the algorithm that cried wolf. "So we have these two challenges," says Liu. "And we need to balance them."

Rather than strike that balance solely for satellite systems, the team endeavored to create a more general framework for [anomaly detection](#)—one that could be applied across industries. They turned to deep-learning systems called generative adversarial networks (GANs), often used for image analysis.

A GAN consists of a pair of neural networks. One network, the "generator," creates fake images, while the second network, the "discriminator," processes images and tries to determine whether they're real images or fake ones produced by the generator. Through many rounds of this process, the generator learns from the discriminator's feedback and becomes adept at creating hyper-realistic fakes. The technique is deemed "unsupervised" learning, since it doesn't require a prelabeled dataset where images come tagged with their subjects. (Large labeled datasets can be hard to come by.)

The team adapted this GAN approach for time series data. "From this training strategy, our model can tell which data points are normal and which are anomalous," says Liu. It does so by checking for discrepancies—possible anomalies—between the real time series and the fake GAN-generated time series. But the team found that GANs alone weren't sufficient for anomaly detection in time series, because they can fall short in pinpointing the real time series segment against which the fake ones should be compared. As a result, "if you use GAN alone, you'll create a lot of false positives," says Veeramachaneni.

To guard against false positives, the team supplemented their GAN with an algorithm called an autoencoder—another technique for unsupervised deep learning. In contrast to GANs' tendency to cry wolf, autoencoders are more prone to miss true anomalies. That's because autoencoders tend to capture too many patterns in the time series, sometimes interpreting an actual anomaly as a harmless fluctuation—a problem called "overfitting." By combining a GAN with an autoencoder, the researchers crafted an anomaly detection system that struck the perfect balance: TadGAN is vigilant, but it doesn't raise too many false alarms.

Standing the test of time series

Plus, TadGAN beat the competition. The traditional approach to time series forecasting, called ARIMA, was developed in the 1970s. "We wanted to see how far we've come, and whether [deep learning](#) models can actually improve on this classical method," says Alnegheimish.

The team ran anomaly detection tests on 11 datasets, pitting ARIMA against TadGAN and seven other methods, including some developed by companies like Amazon and Microsoft. TadGAN outperformed ARIMA in anomaly detection for eight of the 11 datasets. The second-best algorithm, developed by Amazon, only beat ARIMA for six datasets.

Alnegheimish emphasized that their goal was not only to develop a top-notch anomaly detection algorithm, but also to make it widely useable. "We all know that AI suffers from reproducibility issues," she says. The team has made TadGAN's code freely available, and they issue periodic updates. Plus, they developed a benchmarking system for users to compare the performance of different anomaly detection models.

"This benchmark is open source, so someone can go try it out. They can add their own model if they want to," says Alnegheimish. "We want to mitigate the stigma around AI not being reproducible. We want to ensure everything is sound."

Veeramachaneni hopes TadGAN will one day serve a wide variety of industries, not just satellite companies. For example, it could be used to monitor the performance of computer apps that have become central to the modern economy. "To run a lab, I have 30 apps. Zoom, Slack, Github—you name it, I have it," he says. "And I'm relying on them all to work seamlessly and forever." The same goes for millions of users worldwide.

TadGAN could help companies like Zoom monitor time series signals in their data center—like CPU usage or temperature—to help prevent service breaks, which could threaten a company's market share. In future work, the team plans to package TadGAN in a user interface, to help bring state-of-the-art [time](#) series analysis to anyone who needs it.

More information: TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks. arXiv:2009.07769v3 [cs.LG] arxiv.org/abs/2009.07769

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT

research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Method finds hidden warning signals in measurements collected over time (2020, December 17) retrieved 3 May 2024 from <https://techxplore.com/news/2020-12-method-hidden.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.