Dangerous Curves Ahead

Machine Learning Enables Scientists to Spot "Comma-Shaped Clouds," Extreme Weather

Meteorologists can get time-critical help in spotting dangerous cloud formations using artificial intelligence (AI), according to scientists at Penn State and AccuWeather Inc. The team trained an AI with the "machine learning" method, running on PSC's Bridges system, to recognize a typical cloud formation known as comma-shaped clouds satellite images. Their results detected up to 99 percent of the comma-shaped clouds and 64 percent of ensuing stor in 2011 and 2012 satellite images over the U.S. Their hope is to develop an accurate early warning system so storm warnings can be issued more quickly than possible today.

Why It's Important

Between 1980 and 2018, according to the National Oceanic and Atmospheric Administration, 238 particularly sever U.S. weather and climate disasters—including severe storms, floods and other weather problems—caused more than $1.5 trillion in damages. Severe storms alone kill hundreds of people a year. It's safe to say quicker storm warnings would save lives and money.

When meteorologists started using satellite photos to help them predict the weather in the 1970s, they got a surprise Before a storm developed, they saw the clouds in that area took on the shape of a comma. These comma-shaped clouds varied in size from tiny to enormous, with some spanning the continental U.S. from north to south. Before long, though, the meteorologists discovered they had too much of a good thing. The data from satellite photos expanded enormously, quickly overwhelming any ability for humans to monitor them in real time. The big comma-shaped formations were obvious—but smaller ones could still threaten life and property, and were easy to miss.

> "As computer scientists, we were really fascinated by this new field of data science, which we could leverage to make the computers more efficient in solving real-world problems. The issue was to define a problem that had enough data to work with. But not just a lot of data. The data needed to be manually labeled in order for the computer to learn … Meteorology is one of the rare fields where you can find a huge amount of data that are so complex people don't know how to leverage it—and the data are indexed [labeled]."—James Wang, Pennsylvania State University

Enter James Wang of Pennsylvania State University. Wang is a computer scientist who uses AI to solve "Big Data" problems. Around 2008 he was looking for a good scientific application to inspire and test machine-learning technology, in which the computer starts from data that has been labeled by humans, and experiments many, many times on new data until it can reproduce the humans' performance. Detecting severe thunderstorms was a perfect problem. The data are enormous, with tens of thousands of weather stations worldwide and satellites generating gigabytes of high-resolution images, radar data and other data every day. But they have also been carefully reviewed and labeled with "ground truth" by expert meteorologists. Plenty of accurate examples of developing storms were available for the computer to start with.
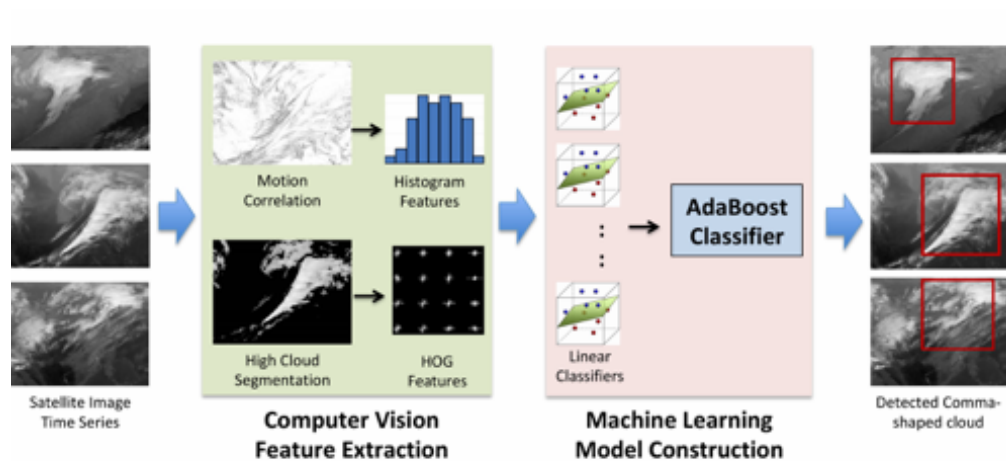
The comma-cloud-detecting AI takes a time series of satellite images (left), first extracting critical features (center left), and then using the AdaBoost classifier to convert those data (center right) into a detection (right).

How PSC Helped

Following up on earlier work by Wang's team, and working with Stephen Wistar and his colleagues at AccuWeather Inc., Wang and his graduate student, Rachel Zheng, started with the satellite images and storm records from the years 2008, 2011 and 2012, because those years had a high number of extreme weather events to study. It was a perfect machine learning dataset. With the help of meteorologists, the team had labeled all the comma-shaped clouds. The storms were a matter of record. But it was also a huge dataset, with more than 50,000 images. To take this problem the team realized they needed a supercomputer with a particular mix of power and compatibility with Big Data. The turned to the Bridges system at PSC. Bridges was designed to be particularly friendly to data science.

Zheng used the first 250 days of 2008 as the training data, with the comma-shaped clouds and storms labeled. That allowed the machine-learning AI to learn how to recognize comma-shaped clouds, as well as to spot signs that a

comma-shaped cloud would develop into a storm. These formations are easy for humans to identify, but they vary ju
enough from cloud to cloud that learning to spot them was a major task for the machine. Because even an AI runnin
on a supercomputer can't process all the weather data from that span of time, Zheng made the AI focus on two facto
One was the shape of a given cloud system. The other was how it changed over time. The two together proved relial
indications of a comma-shaped cloud.

> "The Bridges system works very well with machine learning and image processing. As a result, it took less than 40 seconds for the trained system to process the image data and get a result … Our experience was very good because you have this CPU and large-memory 'multiple choice' that I can choose for a specific job. Even though I process the two steps on different nodes, the data can transfer very quickly."—Rachel Zheng, Pennsylvania State University

Zheng used Bridges' CPU nodes, which are particularly good for the machine learning techniques used in the work.
She also used the system's large-memory nodes, which allowed her to process big image files quickly. After her
program had learned to spot comma-shaped clouds, she used the last 116 days of data from 2008 as a cross-validatic
set, to make sure the learning had worked properly. Finally, she let the program loose on the 2011 and 2012 data,
without any labels, to see how well it predicted the known storms of those years.

> "I still do one operational forecast a week at AccuWeather. We're using all kinds of data, all kinds of screens, taking phone calls from clients … [The machine learning] technique could be useful for calling attention to, say, a comma-shaped disturbance along a cold front in Georgia that none of us saw because we were looking at the snow storm in Michigan. It won't replace meteorologists, but it's a tool that can say, 'You'd better look over here, because something's going on … that four hours from now will be a severe thunderstorm.'"—Stephen Wistar, AccuWeather Inc.

The results were encouraging. The program detected over 99 percent of the comma-shaped clouds correctly. This is
important number. If the machine is going to do a first pass on the data before the humans see it, it can't miss many
them. Equally encouraging, by comparing comma-shaped clouds that did and didn't lead to storms in 2008, the AI
predicted where storms were going to form in the 2011-2012 data more than 64 percent of the time, using satellite
images alone. Best of all, by reviewing the previous 5 hours of weather data, the system could then produce its

predictions in about another 40 seconds. That's fast enough to really help meteorologists spot storm risks, and way faster than humans could review the images. The scientists reported their results in the journal *IEEE Transactions of Geoscience and Remote Sensing* in January 2019. Currently, they're using Bridges' graphics processing units (GPUs) to investigate whether a type of AI called "deep learning," which involves the machine investigating many layers of factors, can improve on their results, particularly when multiple types of weather data are considered together.