

Machine learning looks for useful data in U.S. thunderstorm reports

11 September 2019, by Mike Krapfl



A thunderstorm's shelf cloud approaches the Iowa State campus. Credit: Bill Gallus

Bill Gallus has been known to chase a summer storm or two. But he didn't have to go after this one.

On July 17, 2019, a [thunderstorm](#) approached the Iowa State University campus. Gallus, a professor of geological and atmospheric sciences, headed to the roof above his office in the Agronomy Building. And he didn't forget a camera.

One of his photos shows a shelf cloud marking the edge of severe thunderstorm winds. The cloud's distinct line bisects the photo, low, sharp and imposing, no fluffiness here. The usually busy Osborn Drive outside his office is mostly empty—a few people on the street are turned north-northwest, eyeing the storm.

"The smoothness and low elevation of a shelf cloud makes it an impressive sight to observe," Gallus wrote in a description of the photo. "It forms as the rapidly moving cold air within a

thunderstorm spreads out, lifting the warm humid air quickly above it."

We've all seen dozens of thunderstorms. And the National Weather Service dutifully keeps records of each one and classifies their strength in its Storm Reports database. For a thunderstorm to be marked "severe," for example, it must produce a tornado, hail greater than 1 inch in diameter or winds greater than 58 mph.

But most thunderstorms don't rumble over [wind](#) instruments. So meteorologists have made wind estimates based on storm damage such as trees down, roofs blown away or sheds pushed over. And most of the time, when that kind of wind damage was reported, thunderstorms were simply classified as severe, with no real measurements supporting the designation.

That's a problem for researchers such as Gallus who need good data to help them develop better ways to predict severe, localized thunderstorms.

A big data problem

When Gallus heard campus colleagues from Iowa State's Theoretical and Applied Data Science research group talk about machine learning, he thought the technology's data analysis capabilities could help him study and analyze the Storm Reports database. Maybe the computers could find relationships or connections in the reports that could lead to new forecasting tools?

Well, not so fast, said scientists at the National Oceanic and Atmospheric Administration (NOAA).

The existing severe thunderstorm database maintained by the National Centers for Environmental Information wouldn't be of much use to Gallus or other researchers looking for wind data. The wind reports were unreliable. The reports needed to be cleaned up before they could be

useful for severe wind studies.

So that's what Gallus and a team of Iowa State data scientists are going to do. Supported by a three-year, \$650,000 NOAA grant, they'll use computers and machine learning tools to scour the reports and identify the probability that each one actually describes a thunderstorm with severe winds.

It's no small task—Gallus said the scientists will start with 12 years of severe thunderstorm reports. That's about 180,000 of them.

"And 90 percent of those 180,000 reports contain wind estimates," Gallus said. "They're not based on weather station data. The majority of them say trees or limbs down—somebody called in and said, "My tree blew down.""

Sorting through those reports raises all sorts of challenges for data researchers, said Eric Weber, a project collaborator and Iowa State professor of mathematics.

First, he said the reports are full of data collected by people, not by precise and sophisticated instruments. The reports also contain natural, everyday language. There are idioms, turns of phrases and even typos that have to be analyzed by the machine-learning software.

And second, thunderstorms are very complex. There are many variables—temperature of rising air, condensation, rainfall, lightning and more—that have to be collected, quantified and analyzed to understand the storms.

Weber—who describes machine learning as an artificial neural network that "makes connections based on the information it has available"—said the computer software can handle huge amounts of storm data that would overwhelm teams of people.

Machine-learning software also does that in a very non-human way.

"When we look at data we try to understand the data as human beings," Weber said. "We bring our perceptions and biases. One of the main reasons

machine learning is used so successfully now is that it doesn't bring preconceived notions to its analysis of the data.

"It can find potential relationships that humans can't because of their preconceptions."

Toward better forecasting

As the computers make progress with the storm reports, Gallus said he'll provide updates and demonstrations at NOAA's annual, weeks-long Hazardous Weather Testbed in Norman, Oklahoma. The testbeds are during the May tornado season and are an opportunity for researchers and forecasters to use the latest prediction ideas, tools and technologies.

Gallus hopes to show off the progress of the thunderstorm wind study. He'll collect feedback and suggestions. And all that could eventually lead to a new forecasting tool that predicts the likelihood a thunderstorm will produce severe winds.

"The main need for NOAA right now is to clean up the database for better research," Gallus said. "But we've realized that if this project goes well with [machine learning](#), we could see how it might work as a prediction tool."

Provided by Iowa State University

APA citation: Machine learning looks for useful data in U.S. thunderstorm reports (2019, September 11) retrieved 16 September 2019 from <https://techxplore.com/news/2019-09-machine-thunderstorm.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.