PHYS✷ORG

Hey Siri: How much does this galaxy cluster weigh?

July 19 2022, by Amy Pavlak Laird

Credit: Pixabay/CC0 Public Domain

It's been nearly a century since astronomer Fritz Zwicky first calculated the mass of the Coma Cluster, a dense collection of almost 1,000 galaxies located in the nearby universe. But estimating the mass of

something so huge and dense, not to mention 320 million light-years away, has its share of problems—then and now. Zwicky's initial measurements, and the many made since, are plagued by sources of error that bias the mass higher or lower.

Now, using tools from machine learning, a team led by Carnegie Mellon University physicists has developed a deep-learning method that accurately estimates the mass of the Coma Cluster and effectively mitigates the sources of error.

"People have made mass estimates of the Coma Cluster for many, many years. But by showing that our machine-learning methods are consistent with these previous mass estimates, we are building trust in these new, very powerful methods that are hot in the field of cosmology right now," said Matthew Ho, a fifth-year graduate student in the Department of Physics' McWilliams Center for Cosmology and a member of Carnegie Mellon's NSF AI Planning Institute for Physics of the Future.

Machine-learning methods are used successfully in a variety of fields to find patterns in complex data, but they have only gained a foothold in cosmology research in the last decade. For some researchers in the field, these methods come with a major concern: Since it is difficult to understand the inner workings of a complex machine-learning model, can they be trusted to do what they are designed to do? Ho and his colleagues set out to address these reservations with their latest research, published in *Nature Astronomy*.

To calculate the mass of the Coma Cluster, Zwicky and others used a dynamical mass measurement, in which they studied the motion or velocity of objects orbiting in and around the cluster and then used their understanding of gravity to infer the cluster's mass. But this measurement is susceptible to a variety of errors. Galaxy clusters exist as nodes in a huge web of matter distributed throughout the universe, and

they are constantly colliding and merging with each other, which distorts the velocity profile of the constituent galaxies. And because astronomers are observing the cluster from a great distance, there are a lot of other things in between that can look and act like they are part of the galaxy cluster, which can bias the mass measurement. Recent research has made progress toward quantifying and accounting for the effect of these errors, but machine learning-based methods offer an innovative data-driven approach, according to Ho.

"Our deep-learning method learns from real data what are useful measurements and what are not," Ho said, adding that their method eliminates errors from interloping galaxies (selection effects) and accounts for various galaxy shapes (physical effects). "The usage of these data-driven methods makes our predictions better and automated."

"One of the major shortcomings with standard machine learning approaches is that they usually yield results without any uncertainties," added Associate Professor of Physics Hy Trac, Ho's adviser. "Our method includes robust Bayesian statistics, which allow us to quantify the uncertainty in our results."

Ho and his colleagues developed their novel method by customizing a well-known machine-learning tool called a convolutional neural network, which is a type of deep-learning algorithm used in image recognition. The researchers trained their model by feeding it data from cosmological simulations of the universe. The model learned by looking at the observable characteristics of thousands of galaxy clusters, whose mass is already known. After in-depth analysis of the model's handling of the simulation data, Ho applied it to a real system—the Coma Cluster—whose true mass is not known. Ho's method calculated a mass estimate that is consistent with most of the mass estimates made since the 1980s. This marks the first time this specific machine-learning methodology has been applied to an observational system.

"To build reliability of machine-learning models, it's important to validate the model's predictions on well-studied systems, like Coma," Ho said. "We are currently undertaking a more rigorous, extensive check of our method. The promising results are a strong step toward applying our method on new, unstudied data."

Models such as these are going to be critical moving forward, especially when large-scale spectroscopic surveys, such as the Dark Energy Spectroscopic Instrument, the Vera C. Rubin Observatory and Euclid, start releasing the vast amounts of data they are collecting of the sky.

"Soon we're going to have a petabyte-scale data flow," Ho explained. "That's huge. It's impossible for humans to parse that by hand. As we work on building models that can be robust estimators of things like mass while mitigating sources of error, another important aspect is that they need to be computationally efficient if we're going to process this huge data flow from these new surveys. And that is exactly what we are trying to address—using machine learning to improve our analyses and make them faster."

provided for information purposes only.