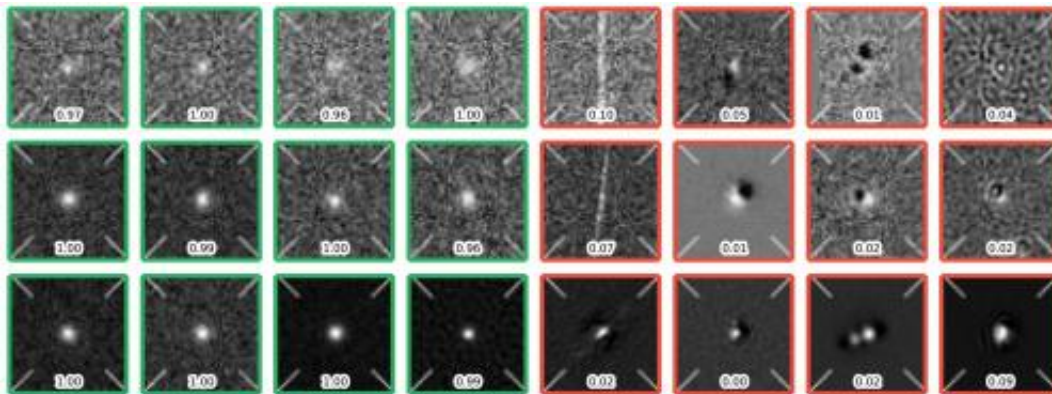


Berkeley algorithms help researchers understand dark energy

November 24 2014, by Linda Vu



UC Berkeley / Berkeley Lab graduate student Danny Goldstein developed a new code using the machine learning technique Random Forest to vet detections of supernova candidates automatically, in real time, optimizing it for the Dark Energy Survey. The technique employs an ensemble of decision trees to automatically ask the types of questions that astronomers would typically consider when classifying supernova candidates. At the end of the process, each detection of a candidate is given a score based on the fraction of decision trees that considered it to have the characteristics of a detection of a supernova. The closer the classification score is to one, the stronger the candidate. Credit: Danny Goldstein, UC Berkeley / Berkeley Lab

Scientists believe that dark energy—the mysterious force that is accelerating cosmic expansion—makes up about 70 percent of the mass and energy of the universe. But because they don't know what it is, they cannot observe it directly.

To unlock the mystery of [dark energy](#) and its influence on the universe, researchers instead must rely on indirect observations—watching how fast cosmic objects, specifically Type Ia supernovae, recede from us as the universe expands. The process of identifying and tracking these objects requires scientists to scrupulously monitor the night sky for slight changes, a task that would be extremely tedious and time-consuming for the Dark Energy Survey (DES) without some novel computational tools developed at the Department of Energy's (DOE's) National Energy Research Scientific Computing Center (NERSC) by researchers at Lawrence Berkeley National Laboratory (Berkeley Lab) and University of California, Berkeley (UC Berkeley).

Currently in its second season, DES snaps pictures of the Southern Sky nightly with DECam, a 570-megapixel camera that is mounted on the Victor M. Blanco telescope at Cerro Tololo Interamerican Observatory (CTIO) in the Chilean Andes. Every night the camera generates between 100 Gigabytes (GB) and 1 Terabyte (TB) of imaging data, which is sent electronically, in real time, to the National Center for Supercomputing Applications (NCSA) and DOE's Fermilab in Illinois for initial processing and archiving.

Object recognition programs developed at NERSC and implemented at NCSA then comb through the images in search of possible detections of Type Ia supernovae. These powerful explosions occur in binary (two-star) systems where one star is a white dwarf. The white dwarf accretes material from a companion star until it reaches a critical mass and explodes in a Type Ia supernova.

"These explosions are remarkable because they can be used as cosmic distance indicators to within 3-10 percent accuracy," says Danny Goldstein, the second-year UC Berkeley graduate student who developed the code to automate the process of supernova discovery on DES images.

Distance is important because the further away an object is located in space, the further back in time it is. By tracking Type Ia supernovae at different distances, researchers can measure cosmic expansion throughout the universe's history. This allows them to put constraints on how fast the universe is expanding and maybe even provide other clues about the nature of dark energy.

"Scientifically, it's a really exciting time, because several groups around the world are trying to precisely measure Type Ia supernovae in order to constrain and understand the dark energy that is driving the accelerated expansion of the universe," says Goldstein, who is also a student researcher in Berkeley Lab's Computational Cosmology Center (C3).

DES begins its search for Type Ia explosions by uncovering changes in the night sky. This is where an image subtraction pipeline developed and implemented by researchers in the DES supernova-working group comes in. The pipeline subtracts images that contain known cosmic objects from new images that are exposed nightly at CTIO. Each night, the pipeline produces between 10,000 and a few hundred thousand detections of supernova candidates that need to be validated.

"Historically, trained astronomers would sit at the computer for hours, look at these dots and offer opinions about whether they had the characteristics of a supernova, or whether they were caused by spurious effects that masquerade as supernovae in the data. This process seems straightforward until you realize that the number of candidates that need to be classified each night is prohibitively large and only one in a few hundred is a real supernova of any type," says Goldstein. "This process is extremely tedious and time-intensive. It also puts a lot of pressure on the supernova working group to process and scan data fast, which is hard work."

So Goldstein developed a code using the machine learning technique

Random Forest to vet detections of supernova candidates automatically, in real time, optimizing it for DES. The technique employs an ensemble of decision trees to automatically ask the types of questions that astronomers would typically consider when classifying supernova candidates. At the end of the process, each detection of a candidate is given a score based on the fraction of decision trees that considered it to have the characteristics of a detection of a supernova. The closer the classification score is to one, the stronger the candidate. Goldstein notes that in preliminary tests, the classification pipeline achieved 96 percent overall accuracy.

"When you do subtraction alone you get far too many 'false-positives'—instrumental or software artifacts that show up as potential supernova candidates—for humans to sift through," says Rollin Thomas, of Berkeley Lab's C3, who was Goldstein's collaborator.

He notes that with the classifier, researchers can quickly and accurately strain out the artifacts from supernova candidates. "This means that instead of having 20 scientists from the supernova working group continually sift through thousands of candidates every night, you can just appoint one person to look at maybe few hundred strong candidates," says Thomas. "This significantly speeds up our workflow and allows us to identify supernovae in real-time, which is crucial for conducting follow up observations."

"Our algorithm can classify a detection of a [supernova](#) candidate in about .01 seconds, whereas an experienced human scanner can take several seconds. Using about 60 cores on a supercomputer we can classify 200,000 detections in about 20 minutes, including time for database interaction and feature extraction," says Goldstein.

Goldstein and Thomas note that the next step in this work is to add a second-level of machine learning to the pipeline to improve the

classification accuracy. This extra layer would take into account how the object was classified in previous observations as it determines the probability that the candidate is "real." The researchers and their colleagues are currently working on different approaches to achieve this capability.

More information: [phys.org/news/2014-08-dark-ene...vey-season-deep.html](https://phys.org/news/2014-08-dark-energy-season-deep.html)

Provided by Lawrence Berkeley National Laboratory

Citation: Berkeley algorithms help researchers understand dark energy (2014, November 24) retrieved 18 April 2024 from <https://phys.org/news/2014-11-berkeley-algorithms-dark-energy.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.