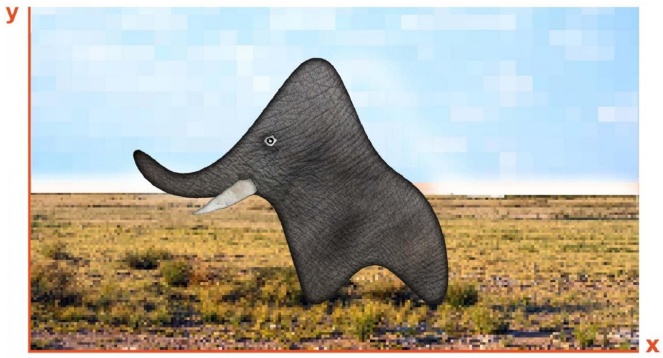
Let's talk about the elephant in the data

3 June 2021



An artistic rendering of how a computer might identify an elephant. Credit: Ben Wigler/CSHL, 2021

You would not be surprised to see an elephant in the savanna or a plate in your kitchen. Based on your prior experiences and knowledge, you know that is where elephants and plates are often to be found. If you saw a mysterious object in your kitchen, how would you figure out what it was? You would rely on your expectations or prior knowledge. Should a computer approach the problem in the same way? The answer may surprise you. Cold Spring Harbor Laboratory Professor Partha Mitra described how he views problems like these in a "Perspective" in *Nature Machine Intelligence*. He hopes his insights will help researchers teach computers how to analyze complex systems more effectively.

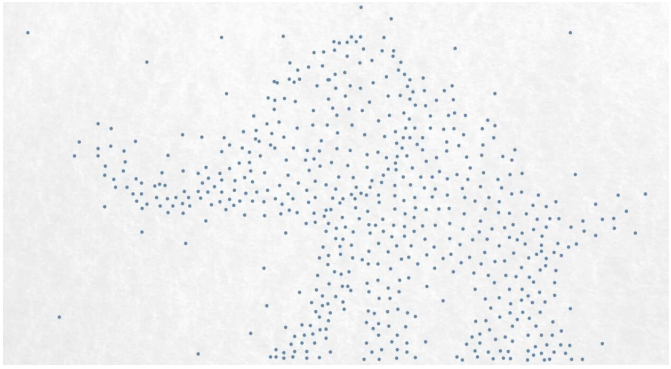
Mitra thinks it helps to understand the nature of knowledge. Mathematically speaking, many data scientists try to create a model that can "fit an elephant," or a set of complex data points. Mitra asks researchers to consider what philosophical framework would work best for a particular machine learning task: "In philosophical terms, the idea is that there are these two extremes. One, you could say "rationalist," and the other, "empiricist" points of view. And really, it's about the role of prior knowledge or prior assumptions."

Rationalists versus empiricists

A rationalist views the world through the lens of prior knowledge. They expect a plate to be in a kitchen and an elephant in a savanna.

An empiricist analyzes the data exactly as it is presented. When they visit the savanna, they no more expect to see an elephant than they do a plate.

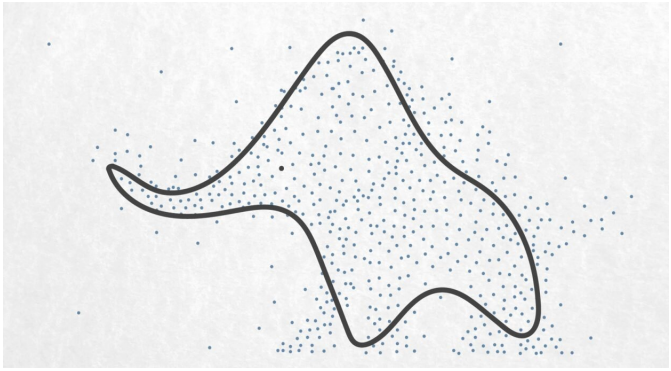
If a rationalist came across this set of data points in the kitchen, they might at first be inclined to view it as a plate. Their prior knowledge states that a plate is likely to be found in a kitchen; it is highly unlikely to find an elephant. They have never seen this situation before, nor have they ever learned that such a situation could occur. Although their result takes in a certain amount of the data, it leaves out other parts. In this case, their methods have produced an incorrect result: a plate.



The ‘data’ in your kitchen. Credit: Ben Wigler

When an empiricist sees the same data, they will analyze it without regard to whether they are in the savanna or their kitchen. They will piece together an image from as many data points as possible. In this case, their result is a jagged image. It doesn't tell the empiricist if they are looking at an elephant, a plate, or anything else.

Neither the empiricist nor the rationalist is wrong. Both approaches work for various kinds of problems. However, in this case, if there is an elephant in the kitchen, it would pay to figure it out as quickly as possible. A middle ground between purely empirical and purely rationalist approaches may be best. With some prior knowledge of what an elephant looks like, you may notice the trunk and legs. And although the chances of an elephant being in your kitchen are low, it is certainly not impossible. Therefore, you would come to the conclusion that there is indeed an elephant in your kitchen, and you probably should leave—fast.



Trunk, legs: must be an elephant! Credit: Ben Wigler

Predictable but wrong

Data scientists face this sort of problem all the time. They train computers to recognize new objects or patterns. Some machine learning programs may be able to process a lot of information and make many rules to fit the presented data, like the jagged image above. The jagged image might be reproducible when the same rules are applied to another similar data set. But just because the pattern is reproducible, that doesn't mean it accurately represents what is happening (the elephant).

There are historical examples of this dilemma. Two thousand years ago, Ptolemy developed a model of the universe that yielded excellent predictions for the movements of the moon and planets. His model was used successfully for centuries. However, Ptolemy used the wrong prior information: He placed the Earth at the center of the solar system and prioritized the circular motions of celestial objects. Johannes Kepler questioned this view in the 17th century and ultimately rejected Ptolemy's approach, which eventually led to Newton's law of universal gravitation. Although Ptolemy's complex model fit his own observations exceptionally well, it did not accurately represent what was happening. Mitra warns that "if you want to be an extreme empiricist, you really do need a lot of data. We now understand why under certain circumstances, such an approach can, in fact, succeed in a mathematically rigorous setting. Biological brains, on the other hand, are halfway in between. You do learn from experience, but you're not entirely data-driven."

Mitra hopes that data scientists will look to brain circuitry for inspiration when developing next-generation machine learning approaches. Vertebrate brains have circuits of different sizes, including medium-sized (mesoscale) ones. Those circuits are encoded with priors (known information, such as what animals look like, where they are found, or how to escape quickly from a charging elephant). At the same time, your brain is highly flexible, classifying new information and weighing the importance of different priors based on experience—elephants may not belong in a kitchen, but somehow, you have one anyway.

Mitra concludes in his article, "This points to the possibility of a new generation of intelligent machinery based on distributed circuit architectures which incorporate stronger priors, possibly drawing upon the mesoscale circuit architecture of vertebrate brains."

More information: Partha P. Mitra, Fitting elephants in modern machine learning by statistically consistent interpolation, *Nature Machine Intelligence* (2021). DOI: 10.1038/s42256-021-00345-8

Provided by Cold Spring Harbor Laboratory

APA citation: Let's talk about the elephant in the data (2021, June 3) retrieved 25 May 2022 from https://techxplore.com/news/2021-06-elephant.html