

Science Finds New Patterns

October 19 2005

People love patterns. Our brains readily organize the stream of information from our senses to find pattern, structure and connections. Our pattern-finding ability lets us understand the world in front of us; sometimes it fools us with optical illusions.

For a growing number of researchers at UC Davis, pattern recognition is a fast-growing field of science where disciplines like computer science, statistics, biology and physics converge. From genetics labs to astronomers' observatories, scientists are being flooded with data. There is a need to develop tools to process data very quickly, so humans can make use of it in real time, and to analyze it for increasingly subtle connections.

"We rely on pattern analysis so much as a society, but haven't stopped to think about its principles," said Nello Cristianini, associate professor of statistics at UC Davis.

Automated pattern recognition has already become a part of our lives. Retailers such as Amazon.com, for example, already use pattern analysis to match customers' demographic data with what they buy, and then use those patterns to recommend additional purchases.

"We need to analyze the human genome, we need to analyze the Web, this is creating a move to unify statistics and computer science, and we call it pattern analysis," said Cristianini, who will co-direct a workshop this fall focusing on those principles, "The Analysis of Patterns," at the Centre "Ettore Majorana" for Scientific Culture in Erice, Italy.

Cristianini's own research includes projects on the evolution of genome sequences and modern languages, working out the evolution of yeast, or the family tree of Indo-European languages, by similarities in DNA sequences or grammar, respectively.

Another of his text-analysis projects, subsumer, applies techniques from DNA analysis to looking for common words in news stories and blog entries on the Internet. Going beyond services like Google News, which automatically collects links to related news stories, subsumer will extract the content from Web pages, recognize significant information such as titles, dates, key words and phrases, and cluster documents in different ways selected by the user.

Given enough data, it is easy to find patterns that do not really exist. Cristianini said he regularly has his students look up their phone numbers in the infinite digits of the number pi. Books have been written about supposed prophecies found in ancient Bible texts. Statistical methods help determine what is truly significant in a mass of data, Cristianini said.

Understanding pattern analysis can also give us insight into how humans -- and machines -- learn from their experiences. The brain of a newborn has to learn to understand the sights, sounds and sensations it encounters, without any instruction from the outside, said Bruno Olshausen, a former professor at the UC Davis Center for Neuroscience and now director of the Redwood Center for Theoretical Neuroscience at UC Berkeley.

Both Olshausen and Cristianini are interested in machine learning -- how to program computers that can learn the way a baby does, by finding structure in a mass of data.

"In unsupervised learning you have to take the data without knowing how you will use it, and organize it by structure," Olshausen said.

Our brains recognize images within fractions of a second -- a task that any modern computer would fail at, although the neural switches in the brain are thousands of times slower than electronic transistors, Olshausen said. The human brain apparently can grasp the "big picture" very quickly, based on the small details of lines and edges.

There is a big, fundamental gap in our understanding of how the brain learns to recognize patterns, he said. Neuroscientists and computer scientists who study machine learning are both pushing against this problem, he said.

"Neuroscience is very well poised for a paradigm shift," Olshausen said.

Pattern recognition helps us pick out the important details and discard the rest. Mathematician Naoki Saito wants to improve that process, both to help humans handle data and to squeeze it down to a manageable size.

As a researcher in the oil and gas industry before joining UC Davis, Saito worked on mathematical theory and algorithms to analyze results from acoustic probes lowered into boreholes. The probes give off sound waves, and the echoes are picked up by a receiver array 10 feet behind the probe tip.

Changes in the rock around the borehole cause changes in the returning sound waves, revealing deposits of natural gas or oil. Those echoes can be extremely complicated, although some experienced oilfield geophysicists can literally "hear" oil deposits, Saito said.

Saito has applied his methods to a wide variety of measurements, such as recordings of brain activity or of the pattern of inks in banknotes.

"You need to sift through the data with mathematical tools to find the patterns that you are interested in," he said. Computers that recognize

patterns as well as we do are "a tall order," Saito said. His work aims to put data from sensors into forms that are easier for humans to use.

Saito is also applying pattern analysis to find better ways to compress large files, such as digital images, without losing crucial information. The popular JPEG file format compresses images by throwing out some data. But this compression always results in some loss of quality, and if the compression goes too far important features get lost.

Saito and Katsu Yamatani of Shizuoka University, Japan, recently filed a patent application for using a technique called "harmonic analysis" to enhance JPEG image compression. Their method breaks down data into two components: smooth, graded or uniform areas, such as skin; and an oscillatory component including stripes, patterns or other abrupt changes. By examining the relationship between the smooth and the oscillatory components the program tries to estimate the information thrown out by compression and reconstruct the image.

In biology, sorting meaningful information from the masses of data generated by genome sequencing and microarray technology has become a major field in just a few years. Stacey Harmer, an assistant professor in the Section of Plant Biology, is using new math tools to find the genes that control a fundamental pattern of the living world -- the daily rhythm of activity by night and day. About 10 percent of all genes in animals and plants show a circadian, or daily, rhythm of activity, she said.

Harmer uses microarrays or "gene chips" to sample the activity of thousands of genes at the same time in the small flowering plant *Arabidopsis thaliana*. By taking measurements at different times over several days, she looks for changes in which genes are switched on or off over 24 hours. By comparing the results to a mathematical wave, she can pick out genes that appear closely tied to circadian rhythms.

Her laboratory has identified a short genetic sequence called the "evening element" that controls genes active at the end of the day. Tinkering with the evening element changes its behavior, leaving evening-regulated genes stuck on or off throughout the day. Other genes known to affect the circadian clock can act through the evening element.

Harmer's group is now looking for similar elements for different phases of the day that interact for 24-hour coverage. They may need more sophisticated mathematical tools to find more subtle effects. This kind of approach in biology will become more and more important, Harmer said.

Jim Crutchfield, a pioneer in chaos and complexity theory, wants to understand where patterns and organization come from, how nature makes them, and how we can discover new ones. His modest goal is to understand how science works: how we make theories from data, and how structure appears at different scales in the universe.

At the Center for Computational Science and Engineering, Crutchfield's computers generate patterns, some complex, some apparently simple, some beautiful and some oddly familiar.

Crutchfield's tools are simple pattern-forming systems called cellular automata. Think of a row of boxes that can be in different states, such as black or white, up or down, one or zero. Write some simple rules for how the boxes behave: for example, if an adjacent box is black, and I am not, turn black. Let them run through 50 or a 100 cycles.

Crutchfield has developed mathematical tools to filter the output from these cellular automata, finding underlying patterns and structures or removing one layer of patterns to expose another. He describes these models as miniature universes, with their own chemistry and physics, which he can use to test his ideas.

Crutchfield's hypothesis is that there are principles of organization that apply to any universe. If we could understand those principles, we could predict what structures would arise in the future starting from a simple set of local rules.

Predicting future behavior of even simple systems is hard, because the possible outcomes soon become chaotic. For example, with modern computer models, forecasters can now make reasonably accurate weather predictions over a few days, but not longer. Throwing more computer power or more raw data at the problem can help, up to a point -- but if we could see the real patterns within the apparent chaos, perhaps we could see further into the future.

"It's time to understand how we build theories. If we can't do that, we're just left with details," Crutchfield said.

Source: UC Davis

Citation: Science Finds New Patterns (2005, October 19) retrieved 18 April 2024 from <https://phys.org/news/2005-10-science-patterns.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.