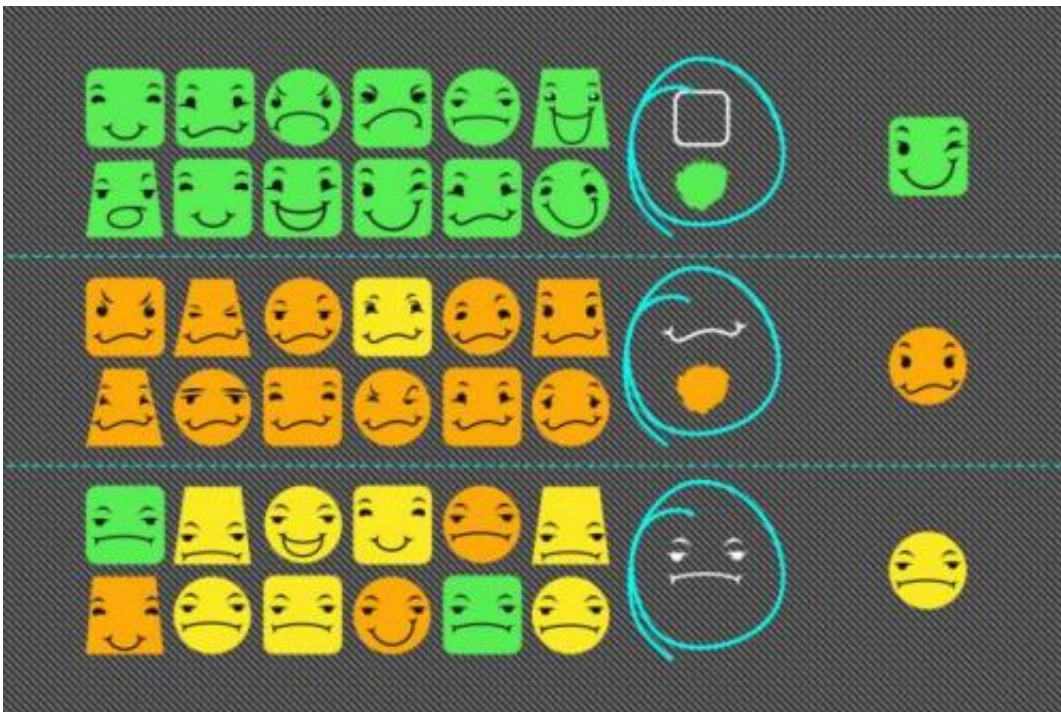# Technique enables pattern-recognition systems to convey what they learn to humans

December 5 2014, by Larry Hardesty

A new machine-learning algorithm clusters data according to both a small number of shared features (circled in blue) and similarity to a representative example (far right). Credit: Christine Daniloff

Computers are good at identifying patterns in huge data sets. Humans, by contrast, are good at inferring patterns from just a few examples.

In a paper appearing at the Neural Information Processing Society's conference next week, MIT researchers present a new system that

bridges these two ways of processing information, so that humans and computers can collaborate to make better decisions.

The system learns to make judgments by crunching data but distills what it learns into simple examples. In experiments, human subjects using the system were more than 20 percent better at classification tasks than those using a similar system based on existing algorithms.

"In this work, we were looking at whether we could augment a machine-learning technique so that it supported people in performing recognition-primed decision-making," says Julie Shah, an assistant professor of aeronautics and astronautics at MIT and a co-author on the new paper. "That's the type of decision-making people do when they make tactical decisions—like in fire crews or field operations. When they're presented with a new scenario, they don't do search the way machines do. They try to match their current scenario with examples from their previous experience, and then they think, 'OK, that worked in a previous scenario,' and they adapt it to the new scenario."

In particular, Shah and her colleagues—her student Been Kim, whose PhD thesis is the basis of the new paper, and Cynthia Rudin, an associate professor of statistics at the MIT Sloan School of Management—were trying to augment a type of machine learning known as "unsupervised."

In supervised machine learning, a computer is fed a slew of training data that's been labeled by humans and tries to find correlations—say, those visual features that occur most frequently in images labeled "car." In unsupervised machine learning, on the other hand, the computer simply looks for commonalities in unstructured data. The result is a set of data clusters whose members are in some way related, but it may not be obvious how.

Balancing act

The most common example of unsupervised machine learning is what's known as topic modeling, in which a system clusters documents together according to their most characteristic words. Since the data is unlabeled, the system can't actually deduce the topics of the documents. But a human reviewing its output would conclude that, for instance, the documents typified by the words "jurisprudence" and "appellate" are legal documents, while those typified by "tonality" and "harmony" are music-theory papers.

The MIT researchers made two major modifications to the type of algorithm commonly used in unsupervised learning. The first is that the clustering was based not only on data items' shared features, but also on their similarity to some representative example, which the researchers dubbed a "prototype."

The other is that rather than simply ranking shared features according to importance, the way a topic-modeling algorithm might, the new algorithm tries to winnow the list of features down to a representative set, which the researchers dubbed a "subspace." To that end, the algorithm imposes a penalty on subspaces that grow too large. So when it's creating its data clusters, it has to balance three sometimes-competing objectives: similarity to prototype, subspace size, and clear demarcations between clusters.

"You have to pick a good prototype to describe a good subspace," Kim explains. "At the same time, you have to pick the right subspace such that the prototype makes sense. So you're doing it all simultaneously."

The researchers' first step was to test their new algorithm on a few classic machine-learning tasks, to make sure that the added constraints didn't impair its performance. They found that on most tasks, it performed as well as its precursor, and on a few, it actually performed better. Shah believes that that could be because the prototype constraint

prevents the algorithm from assembling feature lists that contain internal contradictions.

Suppose, for instance, that an unsupervised-learning algorithm was trying to characterize voters in a population. A plurality of the voters might be registered as Democrats, but a plurality of Republicans may have voted in the last primary. The conventional algorithm might then describe the typical voter as a registered Democrat who voted in the last Republican primary. The prototype constraint makes that kind of result very unlikely, since no single voter would match its characterization.

Road test

Next, the researchers conducted a set of experiments to determine whether prototype-based [machine learning](link) could actually improve human decision-making. Kim culled a set of recipes from an online database in which they had already been assigned categories—such as chili, pasta, and brownies—and distilled them to just their ingredient lists. Then she fed the lists to both a conventional topic-modeling algorithm and the new, prototype-constrained algorithm.

For each category, the new algorithm found a representative example, while the conventional algorithm produced a list of commonly occurring ingredients. Twenty-four subjects were then given 16 new ingredient lists each. Some of the lists were generated by the new algorithm and some by the conventional algorithm, and the assignment was random. With lists produced by the new algorithm, subjects were successful 86 percent of the time, while with lists produced by the conventional algorithm, they were successful 71 percent of the time.

More information: "The Bayesian Case Model: A Generative Approachfor Case-Based Reasoning and Prototype Classification." [interactive.mit.edu/sites/defa … KimRudinShah2014.pdf](link)

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Technique enables pattern-recognition systems to convey what they learn to humans (2014, December 5) retrieved 9 April 2024 from https://phys.org/news/2014-12-technique-enables-pattern-recognition-convey-humans.html