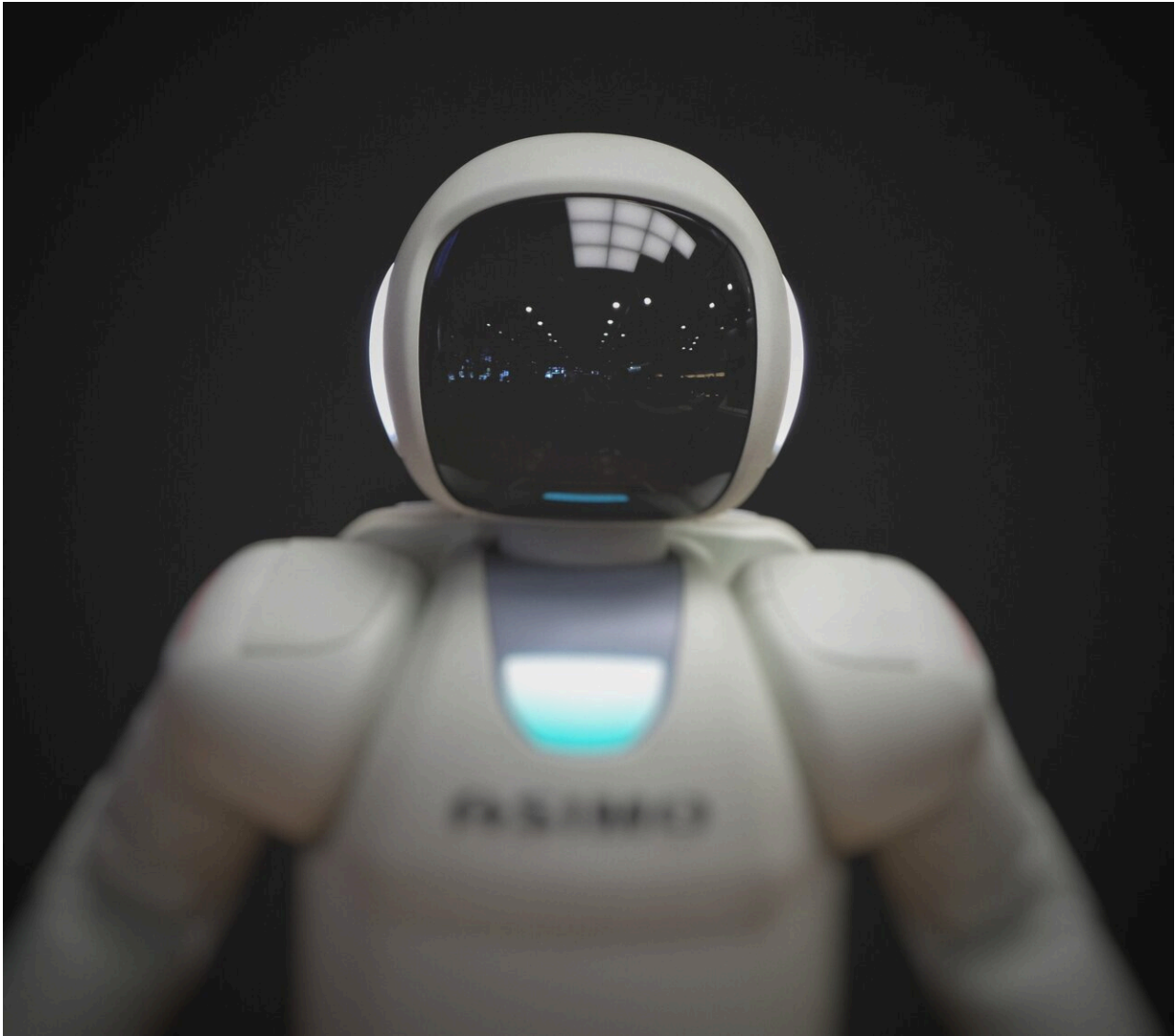


Are machines smarter than a 6-year-old?

March 13 2023, by Julia Cohen



Credit: Unsplash/CC0 Public Domain

Researchers at USC Viterbi's Information Sciences Institute are developing an algorithm that teaches machines to learn without human supervision.

"Generally speaking, machine learning is the science of teaching machines to act similar to humans," said Mohammad Rostami, Research Lead at USC Viterbi's Information Sciences Institute (ISI). Teaching machines to learn without any supervision by humans is the subject of his latest paper, [Overcoming Concept Shift in Domain-Aware Settings through Consolidated Internal Distributions](#), which he presented at the [37th AAAI Conference on Artificial Intelligence](#), held in Washington, D.C. on Feb. 7-14, 2023.

Rostami explained how machine learning is typically done: "We collect data that is annotated by humans, and then we teach the machine how to act similar to humans given that data. The problem we encounter is that the knowledge the machine obtains is limited to the data set that was used for training." Additionally, the data set used for training is often not available after the training process is complete.

The resulting challenge? If the machine receives input that is different enough from the data it was trained on, the machine gets confused and will not act similar to a [human](#).

A bulldog or a shih tzu or something else entirely?

Rostami offered an example, "There are many categories of dogs, different types of dogs are visually not very similar, and the variety is significant. If you train a machine to categorize dogs, its knowledge is limited to the samples that you used for training. If you have a new category of dog that is not among the training samples, the machine is not going to be able to learn that it's a new type of dog."

Interestingly, humans are better at this than machines. When humans are given something to categorize, if they are given just a few samples in a new category (i.e., a new breed of dog), they adjust and learn what that new category is. Rostami said, "a six-year-old child can learn a new category using two, three, or four samples, as opposed to most modern machine learning techniques which require at least several hundred samples to learn that new category."

Categorizing in the face of concept shift

Often, it's not about learning entirely new categories, but being able to adjust as existing categories change.

If a machine learns a category during training, and then over time it undergoes some changes (i.e., the addition of a new subcategory), Rostami hopes that with his research, the machine will be able to learn or extend the notion of what that category is, (i.e., to include the new subcategory).

The changing nature of a category is what is known as "concept shift." The concept of what a category is shifts over time. Rostami offered another real-world example: the spam folder.

He explained, "Your [email service](#) has a model to categorize your inbox emails into legit emails and spam emails. It is trained to identify spam using certain features. For example, if an email is not addressed to you personally, it is more likely that it's spam."

Unfortunately, spammers are aware of these models and constantly add new features in order to trick the models, to prevent their emails from being categorized as spam.

Rostami continued, "this means that the definition of 'spam' changes

over time. It is a time dependent definition. The concept is the same—you have the concept of 'spam'—but over time the definition and details about the concept change. That's concept shift."

A new way to train

In his paper, Rostami has developed a method for training a machine learning model that addresses these issues.

Because original training data is not always available, Rostami's method does not rely on that data. Co-author and ISI Principal Scientist Aram Galstyan explained how, "The model learns the distribution of the old data in the latent space, then it can generate latent representation, almost like generating a synthetic data set by learning the representation of the old data."

Because of this, the model can retain what was learned in the initial training phase, which allows it to adapt and learn new categories and subcategories over time.

It also, importantly, means it will not forget the original [training](#) data or what it learned from it. This is a major issue in [machine learning](#). Galstyan explained, "When you train a new model, it can forget about some patterns that were useful before. This is known as catastrophic forgetting," said Galstyan.

With the approach developed in this paper, Galstyan said "catastrophic forgetting is implicitly addressed because we introduce a correspondence between the old distribution of data and the new one. So, our model will not forget the old one."

What's next?

Rostami and Galstyan are pleased with the results, especially because it does not rely on the availability of source data. Galstyan said, "I was pleasantly surprised to see that the model compares favorably to most of the state-of-the-art existing baselines."

Rostami and Galstyan plan to continue their work on this concept and apply the proposed method on real-world problems.

The paper is also published on the *arXiv* preprint server.

More information: Mohammad Rostami et al, Overcoming Concept Shift in Domain-Aware Settings through Consolidated Internal Distributions, *arXiv* (2022). [DOI: 10.48550/arxiv.2007.00197](https://doi.org/10.48550/arxiv.2007.00197)

Provided by University of Southern California

Citation: Are machines smarter than a 6-year-old? (2023, March 13) retrieved 19 April 2024 from <https://techxplore.com/news/2023-03-machines-smarter-year-old.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.