

Study shows widely used machine learning methods don't work as claimed

March 16 2020



AI will serve to develop a network control system that not only detects and reacts to problems but can also predict and avoid them. Credit: CC0 Public Domain

Models and algorithms for analyzing complex networks are widely used in research and affect society at large through their applications in online social networks, search engines, and recommender systems. According to a new study, however, one widely used algorithmic approach for

modeling these networks is fundamentally flawed, failing to capture important properties of real-world complex networks.

"It's not that these techniques are giving you absolute garbage. They probably have some information in them, but not as much information as many people believe," said C. "Sesh" Seshadhri, associate professor of computer science and engineering in the Baskin School of Engineering at UC Santa Cruz.

Seshadhri is first author of a paper on the new findings published March 2 in *Proceedings of the National Academy of Sciences*. The study evaluated techniques known as "low-dimensional embeddings," which are commonly used as input to [machine learning](#) models. This is an active area of research, with new [embedding](#) methods being developed at a rapid pace. But Seshadhri and his coauthors say all these methods share the same shortcomings.

To explain why, Seshadhri used the example of a social network, a familiar type of complex network. Many companies apply machine learning to social network data to generate predictions about people's behavior, recommendations for users, and so on. Embedding techniques essentially convert a person's position in a social network into a set of coordinates for a point in a geometric space, yielding a list of numbers for each person that can be plugged into an algorithm.

"That's important because something abstract like a person's 'position in a social network' can be converted to a concrete list of numbers. Another important thing is that you want to convert this into a low-dimensional space, so that the list of numbers representing each person is relatively small," Seshadhri explained.

Once this conversion has been done, the system ignores the actual social network and makes predictions based on the relationships between

points in space. For example, if a lot of people close to you in that space are buying a particular product, the system might predict that you are likely to buy the same product.

Seshadhri and his coauthors demonstrated mathematically that significant structural aspects of [complex networks](#) are lost in this embedding process. They also confirmed this result by empirically by testing various embedding techniques on different kinds of complex networks.

"We're not saying that certain specific methods fail. We're saying that any embedding method that gives you a small list of numbers is fundamentally going to fail, because a low-dimensional geometry is just not expressive enough for social networks and other complex networks," Seshadhri said.

A crucial feature of real-world social networks is the density of triangles, or connections between three people.

"Where you have lots of triangles, it means there is a lot of community structure in that part of a social network," Seshadhri said. "Moreover, these triangles are even more significant when you're looking at people who have limited social networks. In a typical social [network](#), some people have tons of connections, but most people don't have a lot of connections."

In their analysis of embedding techniques, the researchers observed that a lot of the social triangles representing community structure are lost in the embedding process. "All of this information seems to disappear, so it's almost like the very thing you wanted to find has been lost when you construct these geometric representations," Seshadhri said.

Low-dimensional embeddings are by no means the only methods being

used to generate predictions and recommendations. They are typically just one of many inputs into a very large and complex machine learning model.

"This model is a huge black box, and a lot of the positive results being reported say that if you include these low-dimensional embeddings, your performance goes up, maybe you get a slight bump. But if you used it by itself, it seems you would be missing a lot," Seshadhri said.

He also noted that new embedding methods are mostly being compared to other embedding methods. Recent empirical work by other researchers, however, shows that different techniques can give better results for [specific tasks](#).

"Let's say you want to predict who's a Republican and who's a Democrat. There are techniques developed specifically for that task which work better than embeddings," he said. "The claim is that these embedding techniques work for many different tasks, and that's why a lot of people have adopted them. It's also very easy to plug them into an existing machine learning system. But for any particular task, it turns out there is always something better you can do."

Given the growing influence of machine learning in our society, Seshadhri said it is important to investigate whether the underlying assumptions behind the models are valid.

"We have all these complicated machines doing things that affect our lives significantly. Our message is just that we need to be more careful about evaluating these techniques," he said. "Especially in this day and age when machine learning is getting more and more complicated, it's important to have some understanding of what can and cannot be done."

More information: C. Seshadhri et al, The impossibility of low-rank

representations for triangle-rich complex networks, *Proceedings of the National Academy of Sciences* (2020). [DOI: 10.1073/pnas.1911030117](https://doi.org/10.1073/pnas.1911030117)

Provided by University of California - Santa Cruz

Citation: Study shows widely used machine learning methods don't work as claimed (2020, March 16) retrieved 25 April 2024 from <https://techxplore.com/news/2020-03-widely-machine-methods-dont.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.