Pressemitteilung

Institute of Science and Technology Austria
Patrick Müller

New Deep Learning Models: Fewer Neurons, More Intelligence

An international research team from TU Wien (Vienna), IST Austria and MIT (USA) has developed a new artificial intelligence system based on the brains of tiny animals, such as threadworms. This novel AI-system can control a vehicle with just a few artificial neurons. The team says that system has decisive advantages over previous deep learning models: It copes much better with noisy input, and, because of its simplicity, its mode of operation can be explained in detail. It does not have to be regarded as a complex "black box", but it can be understood by humans. This new deep learning model has now been published in the journal Nature Machine Intelligence.

Artificial intelligence (AI) can become more efficient and reliable if it is made to mimic biological models. New approaches in AI research are hugely successful in experiments.
Artificial intelligence has arrived in our everyday lives—from search engines to self-driving cars. This has to do with the enormous computing power that has become available in recent years. But new results from AI research now show that simpler, smaller neural networks can be used to solve certain tasks even better, more efficiently, and more reliably than ever before.
An international research team from TU Wien (Vienna), IST Austria and MIT (USA) has developed a new artificial intelligence system based on the brains of tiny animals, such as threadworms. This novel AI-system can control a vehicle with just a few artificial neurons. The team says that system has decisive advantages over previous deep learning models: It copes much better with noisy input, and, because of its simplicity, its mode of operation can be explained in detail. It does not have to be regarded as a complex "black box", but it can be understood by humans. This new deep learning model has now been published in the journal Nature Machine Intelligence.

Learning from nature
Similar to living brains, artificial neural networks consist of many individual cells. When a cell is active, it sends a signal to other cells. All signals received by the next cell are combined to decide whether this cell will become active as well. The way in which one cell influences the activity of the next determines the behavior of the system—these parameters are adjusted in an automatic learning process until the neural network can solve a specific task.
"For years, we have been investigating what we can learn from nature to improve deep learning," says Prof. Radu Grosu, head of the research group "Cyber-Physical Systems" at TU Wien. "The nematode C. elegans, for example, lives its life with an amazingly small number of neurons, and still shows interesting behavioral patterns. This is due to the efficient and harmonious way the nematode's nervous system processes information."
"Nature shows us that there is still lots of room for improvement," says Prof. Daniela Rus, director of MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL). "Therefore, our goal was to massively reduce complexity and enhance interpretability of neural network models."
"Inspired by nature, we developed new mathematical models of neurons and synapses," says Prof. Thomas Henzinger, president of IST Austria.
"The processing of the signals within the individual cells follows different mathematical principles than previous deep learning models," says Dr. Ramin Hasani, postdoctoral associate at the Institute of Computer Engineering, TU Wien and MIT CSAIL. "Also, our networks are highly sparse—this means that not every cell is connected to every other cell. This also makes the network simpler."

Autonomous Lane Keeping

To test the new ideas, the team chose a particularly important test task: self-driving cars staying in their lane. The neural network receives camera images of the road as input and is to decide automatically whether to steer to the right or left.

"Today, deep learning models with many millions of parameters are often used for learning complex tasks such as autonomous driving," says Mathias Lechner, TU Wien alumnus and PhD student at IST Austria. "However, our new approach enables us to reduce the size of the networks by two orders of magnitude. Our systems only use 75,000 trainable parameters."

Alexander Amini, PhD student at MIT CSAIL explains that the new system consists of two parts: The camera input is first processed by a so-called convolutional neural network, which only perceives the visual data to extract structural features from incoming pixels. This network decides which parts of the camera image are interesting and important, and then passes signals to the crucial part of the network – a "control system" that then steers the vehicle.

Both subsystems are stacked together and are trained simultaneously. Many hours of traffic videos of human driving in the greater Boston area were collected, and are fed into the network, together with information on how to steer the car in any given situation—until the system has learned to automatically connect images with the appropriate steering direction and can independently handle new situations.

The control part of the system (called neural circuit policy, or NCP), which translates the data from the perception module into a steering command, only consists of 19 neurons. Mathias Lechner explains that NCPs are up to 3 orders of magnitude smaller than what would have been possible with previous state-of-the-art models.

Causality and Interpretability

The new deep learning model was tested on a real autonomous vehicle. "Our model allows us to investigate what the network focuses its attention on while driving. Our networks focus on very specific parts of the camera picture: The curbside and the horizon. This behavior is highly desirable, and it is unique among artificial intelligence systems," says Ramin Hasani. "Moreover, we saw that the role of every single cell at any driving decision can be identified. We can understand the function of individual cells and their behavior. Achieving this degree of interpretability is impossible for larger deep learning models."

Robustness

"To test how robust NCPs are compared to previous deep models, we perturbed the input images and evaluated how well the agents can deal with the noise," says Mathias Lechner. "While this became an insurmountable problem for other deep neural networks, our NCPs demonstrated strong resistance to input artifacts. This attribute is a direct consequence of the novel neural model and the architecture."

"Interpretability and robustness are the two major advantages of our new model," says Ramin Hasani. "But there is more: Using our new methods, we can also reduce training time and the possibility to implement AI in relatively simple systems. Our NCPs enable imitation learning in a wide range of possible applications, from automated work in warehouses to robot locomotion. The new findings open up important new perspectives for the AI community: The principles of computation in biological nervous systems can become a great resource for creating high-performance interpretable AI—as an alternative to the black-box machine learning systems we have used so far."

Video: https://youtu.be/8KBOf7NJh4Y

Code Repository: https://github.com/mlech26l/keras-ncp

Originalpublikation:

https://www.nature.com/articles/s42256-020-00237-3

The team tested their new neural network on an important task: keeping self-driving cars in their lanes.
Ramin Hasani
Ramin Hasani