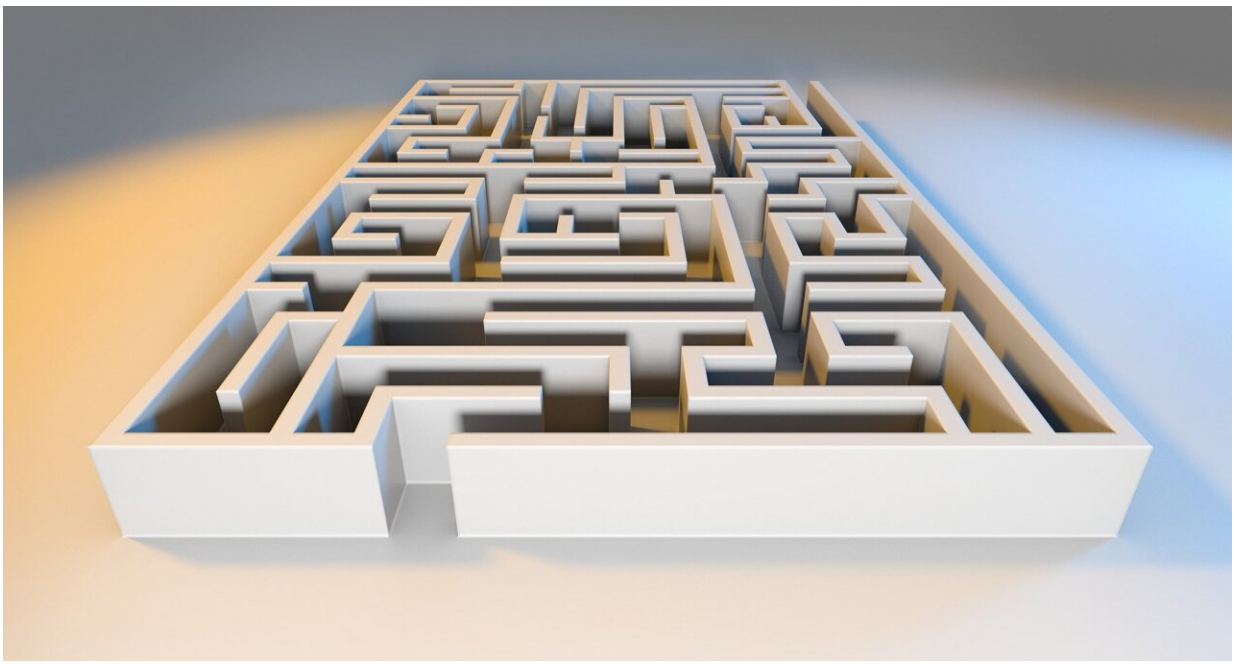
Machine learning of binary 'yes/no' systems may improve medical diagnoses, financial risk analysis, and more

November 11 2022



Credit: Pixabay/CC0 Public Domain

Similar to a mouse racing through a maze, making "yes" or "no" decisions at every intersection, researchers have developed a way for machines to swiftly learn all the twists and turns in a complex data system.

"Our method may help improve the diagnosis of urinary diseases, the imaging of cardiac conditions and analysis of financial risks," reported Abd-AlRahman Rasheed AlMomani of Embry-Riddle Aeronautical University's Prescott, Arizona, campus.

The research was accepted for the Nov. 11 edition of the journal *Patterns* with Jie Sun and Erik Bollt of Clarkson University's Center for Complex Systems Science. The goal of the work is to more efficiently analyze binary ("Boolean") data.

"We can see everything around us as a network of objects and variables that interact with each other," said AlMomani, assistant professor of Data Science and Mathematics at Embry-Riddle. "Understanding those interactions can improve our predictions and management of a whole host of networks—from biology and gene regulatory networks, to even air flight."

Boolean, or "yes/no," data are frequently used in the field of genetics, where gene states may be described as "on" (with high gene expression), or "off" (with little or no gene expression), AlMomani explained. Learning Boolean functions and networks based on noisy observational data is key to deciphering many different science and engineering problems—from plant-pollinator dynamics and drug targeting, to assessing a person's risk of tuberculosis.

The challenge, AlMomani explained, is that the standard method for learning Boolean networks—called REVEAL (for reverse engineering algorithm for interference of genetic network architectures)—blends many different sources of information. The REVEAL approach thus increases computational complexity and costs, and researchers must dampen noise to analyze all of the data. Further, the REVEAL method isn't optimal for solving quantitative biology problems, which require uncovering causal factors.

To weed out incorrect answers faster, AlMomani and colleagues leveraged a method called Boolean optimal causation entropy, which progressively narrows the number of correct solutions to a problem. The method essentially turns a complex diagnostic process into a decision tree, where yes/no questions such as, "Does the patient have a fever? Nausea? Lumbar pain?" can guide a clinician to the correct diagnosis.

AlMomani explained that many different scientific questions hinge upon "a Boolean variable that is basically zero or one. An event happened or it did not happen. A patient will have a test and get a positive or a negative result. We can then categorize that patient's test results, health history and outcomes as Boolean variables."

To test their ideas, the researchers got their hands on a complete set of 958 possible board configurations at the end of a Tic-Tac-Toe game. The board and different game moves were then expressed as mathematical problems in order to predict which player would win.

The researchers also tested their method using a dataset from cardiac spectroscopy images. Their system got the diagnosis right 80% of the time.

The *Patterns* article is titled "Data-Driving Learning of Boolean Networks and Functions by Optimal Causation Entropy Principle (BoCSE)."

More information: Abd AlRahman R AlMomani, Data-Driven Learning of Boolean Networks and Functions by Optimal Causation Entropy Principle (BoCSE), *Patterns* (2022). DOI: 10.1016/j.patter.2022.100631. www.cell.com/patterns/fulltext … 2666-3899(22)00263-X

TechXplore

Provided by Embry-Riddle Aeronautical University

Citation: Machine learning of binary 'yes/no' systems may improve medical diagnoses, financial risk analysis, and more (2022, November 11) retrieved 26 April 2024 from https://techxplore.com/news/2022-11-machine-binary-yesno-medical-financial.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.