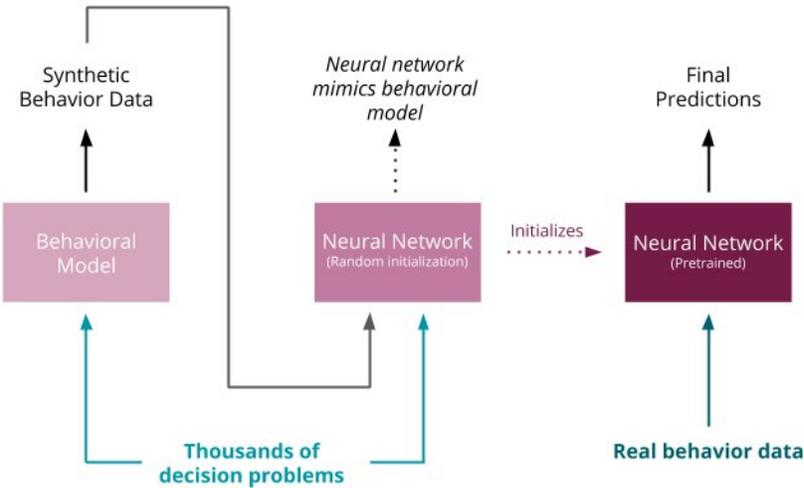


Infusing machine learning models with inductive biases to capture human behavior

June 7 2019, by Ingrid Fadelli



A flowchart summarizing the researchers' approach. First, they generated synthetic behavior data by applying a behavioral model from psychology to a large collection of decision problems. They then trained a neural network to predict this synthetic behavior, effectively transferring the behavioral model into the network. Once the network finished learning the synthetic data they fine tuned it on real human data, allowing it to further build upon the cognitive model and achieve better predictions. Credit: Bourgin et al.

Human decision-making is often difficult to predict and delineate theoretically. Nonetheless, in recent decades, several researchers have developed theoretical models aimed at explaining decision-making, as well as machine learning (ML) models that try to predict human behavior. Despite the achievements associated with some of these models, accurately predicting human decisions remains a significant research challenge.

ML techniques might seem ideal for tackling decision-making prediction problems, yet it is still unclear whether they can actually improve predictions made by theoretical models. Researchers at the University of California (UC) Berkeley and Princeton University have recently carried out a study exploring the effectiveness of ML in capturing [human behavior](#). In their paper, set to be presented at the International Conference on Machine Learning and [pre-published on arXiv](#), they propose a new approach for predicting human decisions, which they refer to as 'cognitive [model](#) priors.'

"ML has revolutionized our ability to predict phenomena in a number scientific domains," David Bourgin, one of the researchers who carried out the study, told TechXplore. "In psychology and economics, however, ML approaches for prediction purposes are still relatively rare. One reason for this is that many off-the-shelf ML models require a significant amount of data to train, and behavioral datasets tend to be fairly small."

In machine learning studies, the standard method of tackling issues related to small datasets is to restrict the space of possible solutions. However, this is not always a straight-forward task, particularly when working with neural networks, as a sufficiently general and easily applicable method for dealing with small datasets does not yet exist.

"We were motivated by the idea that we might improve the extent to

which we could predict certain behavioral phenomena if we could somehow translate insights from psychological theories into inductive biases within a machine learning model," Bourgin said.

The study carried out by Bourgin and his colleagues made two significant contributions to the study of ML for human decision-making prediction. Firstly, the researchers introduced the concept of 'cognitive model priors,' which entails pre-training neural networks with synthetic data derived using established [theoretical models](#) developed by cognitive psychologists. This approach allowed them to also introduce the very first large-scale [dataset](#) for training algorithms on human decision-making tasks.

"Our approach combines existing scientific theories of human behavior with the flexibility of [neural networks](#) to adapt to best predict human risky monetary decisions," Joshua Peterson, another researcher involved in the study, told TechXplore. "We do this by converting a behavioral model into a more flexible form by training a neural network to approximate it. After this step, the neural network will already be nearly as predictive as the behavioral model, and is now in a place to make the most of further learning from real examples of human behavior."

Using 'cognitive model priors' the researchers attained state-of-the-art results on two existing benchmark datasets. These findings suggest that it is indeed possible for ML models to make accurate decision-making predictions, even if available datasets are small. In their case, this was achieved by pre-training models on artificial data derived from cognitive models.

"Our key theoretical contribution is the introduction of a general way to translate between psychological models and machine learning approaches," Bourgin said. "The upshot is that this can help researchers to apply machine learning models to behavioral datasets that would

otherwise be too small. We hope that this will encourage greater collaboration between the machine learning and behavioral science communities by providing a way to evaluate a wider class of models of human decision making."

In their study, Bourgin, Peterson and their colleagues have made significant advancements in the study of ML tools for capturing human behavior, with their approach achieving unprecedented performance on two restricted datasets of human decisions. They also presented a new dataset that contains 240,000 human judgments across 13,000 decision problems, which could be used by other research groups to train their own ML models. From a practical standpoint, their work could save researchers the significant amount of time that is typically spent on collecting data for ML human prediction models.

"We're excited to see what other domains of human behavior could benefit from our approach, especially in more natural settings," Peterson said. "We're also interested to find ways to close the loop by using the improved [machine learning](#) models to discover new scientific theories."

More information: David D. Bourgin et al. Cognitive model priors for predicting human decisions. arXiv:1905.09397v1 [cs.LG].arxiv.org/abs/1905.09397

© 2019 Science X Network

Citation: Infusing machine learning models with inductive biases to capture human behavior (2019, June 7) retrieved 26 April 2024 from <https://techxplore.com/news/2019-06-infusing-machine-inductive-biases-capture.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.