Adv Robot Autom 2018, Volume: 7
DOI: 10.4172/2168-9695-C1-015

How to account for the unknown when training supervised machine learning algorithms?

Thomas Hahn
University of Arkansas at Little Rock, USA

Incomplete feature selection causes reproducible prediction errors. Such prediction errors are the consequences of still Imperatively Hidden Objects (IHO). They are called IHO because – even though they cannot be directly observed or measured – they affect the predicted outcome. Such IHOs are often the hidden causes. Their consequences are exposed by reproducible prediction errors. IHOs are challenging because their exact impact on the observation of interest cannot be fully considered until they have been fully uncovered. But, since feature selection must be completed before any supervised machine learning (ML) algorithm can be properly trained, such kind of IHOs must be discovered before feature selection can be fully completed. To distinguish IHOs from one another and from background noise, the measuring methods and their surrounding experimental environment must be varied. Independent features can be considered as singlets, e.g., the impact of transcription, length of poly-(A)-tails and ribosomal coverage on protein abundance, because they don't depend on one another. However, codons must be considered as atomic triplets only. When the prediction can be made by different methods in different dimensions, feature selection can be considered as fully completed, e.g., when transcription can be fully predicted either by considering the trajectories of time series plots or Transcription Factor Binding Site (TFBS) distributions, Transcription Factor (TF) ratios and TF abundance. New features can be discovered by creating external experimental conditions, which cause any until then correctly ML prediction algorithm to fail, because it indicates another still unknown dimension by which IHO differ from one another and their background noise environment. Strategies for uncovering IHOs are discussed. Focusing on IHO discovery can speed up our scientific progress because each newly uncovered IHO should be added to the selected features for training new ML algorithms because IHOs prevents better understanding of complex phenomena, such as aging.

Hahn5Thomas@gmail.com