AI and Crop Improvement

true

June 2022

Abstract

Lorem ipsum dolor sit amet.

Contents

Crop Improvement

To achieve a rate of yield increase that keeps pace with the growing world population, new approaches to crop improvement will be required.

Modern Crop Improvement

Conventional breeding took advantage of traditional phenotyping methods to inform crossing and selection.

Prior issues facing crop improvement research revolved around data accumulation. Genotyping methods prior to NGS were laborious and limited quantities of genotypes were achievable. Traditional phenotyping methods are time intensive and sometimes limited by the expertise of evaluators. Envirotyping was difficult to achieve with useful levels of resolution. The post-NGS sequencing technology era has made sequencing whole genomes a feasible genotyping approach. image-based phenotyping and other high throughput techniques generate terabytes of data every growing season. Remote sensing technologies have made observing envirotypes within a single field possible.

With the increase in genetic technology availability and decrease in cost during the (1980s-2000s), targeted improvements through approaches such as marker assisted selection became feasible. Genetic improvement was seen as the new way forward in contrast to the plateauing gains made from cultural practices. Genetic data became abundant in the next-generation sequencing era. Earlier QTL studies have identified many loci with large contributions to highly heritable traits. Traits with unresolved genetic architecture, result from the contributions of many, small effect loci. QTL studies in such traits require extensive phenotyping and are often not stable across environment.

Modern crop improvement has the advantage and challenge of access to large genomic, phenomic and enviromic data sets to inform breeding and selection decisions. A single population can now produce dramatically more data then was cumulatively obtained in the breeding process in previous decades. The prior issues of genotyping and phenotyping bottlenecks have been largely overcome with higher levels of automation achieved through approaches like post-NGS sequencing technologies , HTP, and remote sensing . However, these solutions and the accompanying "big data" also lend new issues to the field of crop improvement. Fully exploiting "-omic" data in an efficient manor is difficult with prior approaches. There is now a need for new methods of analysis which effective and efficient in order to continue bridging the gap between observed genotypes and observed phenotypes. AI has the potential to provide diverse solutions to these ends.

AI

Artificial intelligence (AI) is concerned with the process of designing computers that can think and act humanly and rationally (Russell and Norvig 2009). In recent years, AI has been increasingly explored as a means to analyze big data most popularly through machine learning (ML) approaches. However in addition to ML, AI encompasses a number of diverse sub-fields which can can be generally divided into knowledge-base, statistics-based and [subsymbolic] systems.

Algorithms that do not attempt learning are considered to be rules-based. Rules-based systems are generally implemented by defining a series of situation-action (i.e. if-then) rules. Defining these rules requires previous knowledge of the relationship between input-output pairs(Hayes-Roth 1985). Programming a system that can anticipate all situation-actions which lead from input to output variables is more difficult than training a system that can learn from input (Jordan and Mitchell 2015). At the core of modern AI is machine learning (ML). Machine learning is a subfield within AI which attempts *learning* or improvement through experience (Russell and Norvig 2009; Libbrecht and Noble 2015; Liakos et al. 2018). ML can be characterized by feedback type, data type . . . .

Machine learning can be considered to be supervised or unsupervised given the system's access to feedback. Supervised learning occurs in two phases, training and testing. During the training phase, predetermined input-response pairs (labeled data) are used as examples and the learning algorithm attempts to formulate function that connect input data to respective labels (Liakos et al. 2018; Montesinos López et al. 2022, p. 29). During the testing phase of supervised learning, the learned pattern (trained model) is used to generate label predictions and the accuracy of the predictions can be evaluated against user-defined labels. In this type of system the feedback is considered to be explicit. In unsupervised learning, the absence of preassigned labels for the input data only allows the model to evaluate patterns and prevents prediction accuracy from being evaluated since no correct input-response pairs have been specified. This type of system has no feedback. Systems that utilize both supervised and unsupervised learning are termed semi-supervised.

Machine learning models can also be divided by learning type. Within supervised learning

Machine learning exists at the intersection of computer science and statistics (Fig. 1) and thus many statistical models are considered ML when applied with a prediction-centric approach. The field of statistics by comparison puts a greater focus on inference over prediction (Bzdok et al. 2018), but methods shared by the two fields can generally accomplish both. For this reason

AI in Crop Improvements

Inductive Logic Programming

Robotic Process Automation

Expert Systems/Fuzzy Systems

References

Bzdok D, Altman N, Krzywinski M (2018) Statistics versus machine learning. Nature Methods 15:233–234. https://doi.org/10.1038/nmeth.4642

Hayes-Roth F (1985) Rule-based systems. Communications of the ACM 28:921–932

Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. Science 349:255–60. https://doi.org/10.1126/science.aaa8415

Liakos K, Busato P, Moshou D, et al (2018) Machine Learning in Agriculture: A Review. Sensors 18:2674. https://doi.org/10.3390/s18082674

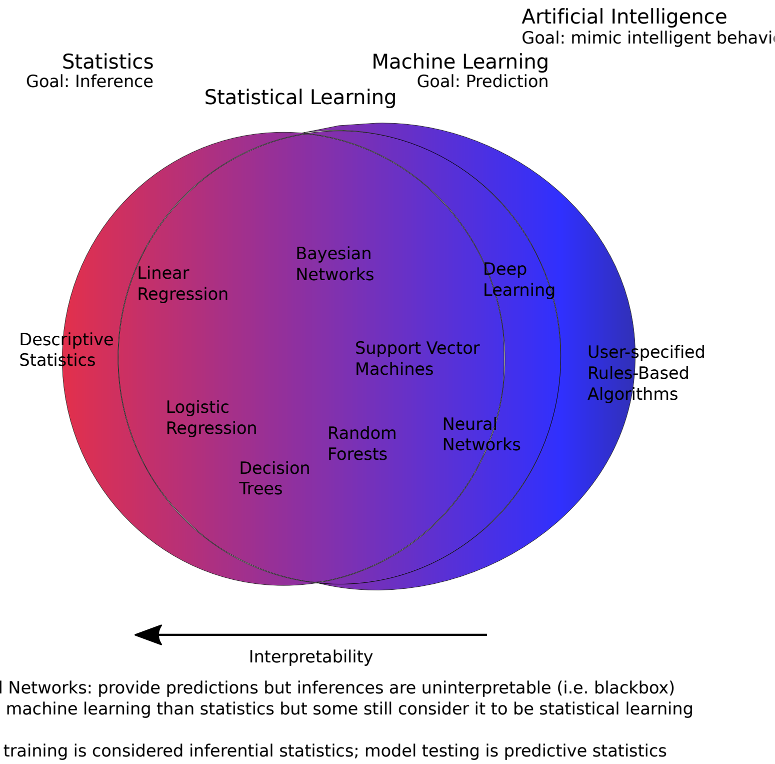

Figure 1: Caption.

Libbrecht MW, Noble WS (2015) Machine learning applications in genetics and genomics. Nature Reviews Genetics 16:321–332. https://doi.org/10.1038/nrg3920

Montesinos López OA, Montesinos López A, Crossa J (2022) Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer International Publishing, Cham

Russell S, Norvig P (2009) Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall Press, USA