Thesis Portfolio

Trait-Predicting Algorithms: Proposed Algorithm for Predicting Phenotypes from Genetic Data

(Technical Report)

Social and Ethical Attitudes Towards Trait Prediction via Genomic Samples

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Paul Vann

Spring, 2023

Department of Computer Science

Table of Contents

Executive Summary

In the STS Research Paper, the ethical dilemmas associated with trait predicting algorithms are evaluated. This evaluation utilizes a review of critical literature, an analysis of existing legislation on genetic privacy policies, and a poll of UVA students on their opinions on trait-predicting algorithms to establish a conclusion about how trait prediction algorithms should be handled in the future. While genetic trait prediction algorithms provide significant benefit to areas such as law enforcement and analysis of centuries old DNA, there are the ethical issues of privacy and racial bias that present an issue. Because trait predicting algorithms utilize genetic data in their training and testing, the issue of privacy and how that data can be effectively secured becomes an issue. Furthermore, when used for tasks such as identifying a suspect in a crime, these algorithms can show negative bias towards specific racial groups if they are not trained and managed correctly. Using the STS Frameworks Technological Determinism and Social Construction of Technology, conclusions were able to be drawn about how trait predicting algorithms should be handled in the future. Through the results from the poll of UVA students as well as the literature review on genetic algorithms and how they have benefited society in the past, it is concluded that these algorithms should be able to exist and be used for good. However, through the evaluation of existing legislation and regulations on genetic privacy and the review of critical literature on racial bias in genetic algorithms, in order for these algorithms to be ethical there must be transparency of the algorithm and regulations on how they are trained.

In the technical paper, a proposed algorithm for a trait prediction algorithm is presented. In the past, trait prediction has been done primarily manually and only applied to a small number of phenotypes. This proposed algorithm utilizes a neural network structure to identify much more complex phenotypes, by finding patterns in genetic data using “neurons”. All aspects of the

training and construction of the algorithm are covered in the paper. The paper proposes an ideal training set including what features or traits to include, as well as how to structure this data into machine readable data. Furthermore, the most ideal system for training this algorithm was determined through an analysis of cheap and easy to use machine learning GPU's. The ideal structure for the algorithm was determined based on the inputs, the task at hand, and the accuracy necessary for the algorithm to be viable. While I did not have genetic data to test this algorithm or create an instance of it, the expected results of this concept are extensive and include the ability for researchers who have such data to create a highly accurate trait predicting algorithm. Overall, the goal of this conceptual algorithm is to present unique research on a neural network based trait prediction algorithm, and allow researchers in the future to develop a tangible algorithm that can be used for good.

The combination of these two papers creates a big picture of how to build a trait prediction algorithm, while also keeping in mind the unethical means in which such an algorithm could be designed or used. By referencing the STS paper when designing such an algorithm, researchers can take into consideration how to prevent racial bias and secure the data used within the algorithm much more efficiently. Furthermore, these two papers represent as a whole the moral and ethical dilemmas associated with genetic algorithms, and while they can have a benefit, the numerous issues associated with them.