The Upgrade Proposal

Theo Portlock

December 6, 2020

Abstract

In our lives, we try to make decisions that increase the possiblity of a moral outcome. We judge others by their moral objectives and their adherance to them. Identifying, understanding, and mitigating the flaws of human thought process and decision making is essential for human progress. The aim of this project is to create a tool that will make better, more informed moral decisions than a human. The possibility of succeeding is incalcuable but increases with effort, time, and expertise. However, if good progress towards these aims is made, the consequences will pave the way for the future of a more moral society.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Morality

Morality is the set of priciples used by individuals that define right and wrong actions. A moral agent is an individual that acts within a moral system. An event is concidered moraly virtuous (good) with respect to a moral system if its consequences increase the probability of a moral end. An event is concidered moraly reprehensible (bad) with respect to a moral system if its consequences decrease the probability of a moral end. A moral system can contain multiple ends but often, a combination of these ends weighted by personal importance can be assumed to be an end in of itself.

1.1.1 Majority moral opinion

Moral ends vary between individuals. When individual moral agents interact, one of three possible combinations can be observed. Firstly, a mutual rejection of moral systems can be defined as the assumption of one anothers actions are bad. Secondly, an unbalanced interaction is a consequence of the willingness of only one moral agent to encourage the others persuit of their respective moral system. Lasty, a mutual acceptance of moral systems is the understanding that the ends described by a moral system is aligned to the point that their continued actions weighted by their estimated probability is similar to those actions taken by the moral system of another individual. As with the legal system, an estimation of common "good" is necessary for collective arbritration of an action. Once a consensus is established, the application of the golden rule, the principle of treating others as you wish to be treated, forms the idea of justice and social contract. The search for this consensus is done at a personal scale during adolesence, or at a collective scale when we vote, polls, or protest. Communication of intention can often be reciprocal, that is to say one individual has and understanding of the others understanding of your understanding and so on. An understanding of the consequences of breaking this rule encourages the enforcement of a social contract

1.1.2 How morality changes

Consensus morality changes over time. Moral opinions can be changed with varying probabilities. The only means of moral re-evaluation is through the revelation of inconsistencies. This can be due to incomplete data relating to a moral question or cognitive dissonance.

1.2 Progress

1.2.1 How is society improving?

1.2.2 How is society deteriorating?

1.3 The fallibility of human decision making

1.3.1 Bias

1.3.2 Misinformation

1.3.3 Missing information

1.4 Alternative decision making algorithms

1.4.1 Deep learning

1.4.2 HTM

Numenta (Numenta, n.d.), alpha go, etc

1.4.3 How to align intelligent systems with consensus moral beliefs

1.5 AI saftey and responsibility

1.5.1 The ”off button” problem

1.5.2 The paperclip problem

1.6 Project aims

Chapter 2

Datastreams

2.1 Sources of data

2.2 Binary conversion

2.3 Sparcity

2.4 Tandem data input

Chapter 3

Combinations

3.1 Classical combinatorics

3.2 The combination problem of scale

3.3 The combinations array

3.4 Slicing the combinations array

3.4.1 Combinations of combinations

Chapter 4

Memory

4.1 Persistance of activation

4.2 Edge contraction

4.3 Delay function

4.4 Transfer and Storage

4.5 Mind meld

Chapter 5

Action

5.1 How do machines act

5.2 The random-pianist method

Chapter 6

Runtime

6.1 Chapter introduction

Chapter 7

Testing

7.1 Chapter introduction

This chapter will focus on some of the more well known problems that face current machine learning models.

7.2 Test 1 - Mathematics

7.2.1 Addition

7.3 Test 2 - Natural language modelling

7.3.1 Text prediction

7.4 Test 3 - Signal processing

7.4.1 Voice recognition

7.5 Test 4 - Image classification

7.5.1 Medical diagnosis

References

Numenta, N. H. (n.d.). *Home page, numenta inc, 2007.*

Acknowledgements

Appendix

7.6 Chapter introduction