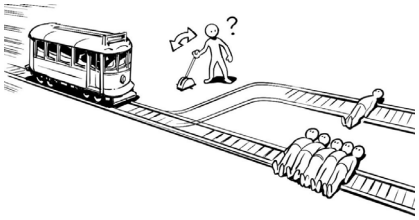
Moral Mining

Let Machines learn the Concept of Moral Values and Ethics



Abstract

While discussing ethics in the context of Data Engineering, the problem of biased algorithms is omnipresent. Seemingly morally neutral decisions are influenced by programmers. These biases influence individuals and their decision making rather than letting the algorithm act on the basis of the individual users moral values. The need for computers to understand moral values and therefore ethics seems inevitable.

With the help of Machine Learning, morality can be formally described and applied to individuals and their personal moral values. As moral values depend on each other and are highly correlated, the application of a thesaurus is possible to define correlations in ethics.

Problem

Biases and moral values of programmers lead to algorithms that have biases as well. This can affect each individual in a way that one has to deal with propositions and recommendations that are against one‘s own moral values. Therefore, it would be beneficial if these effects could be according to ones personal moral values and biases.

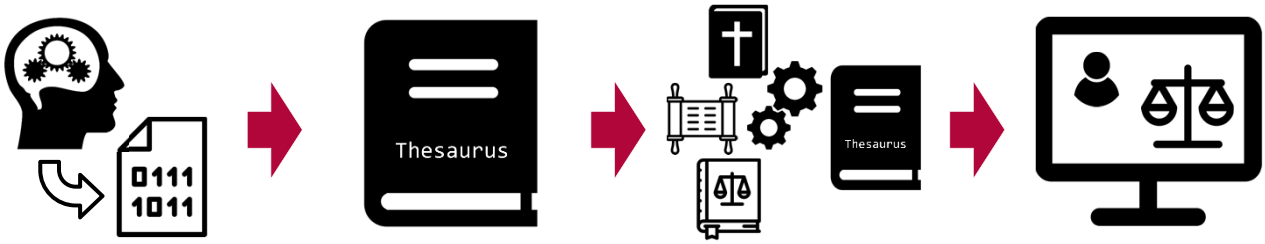
One would prefer recommendations according to one‘s own moral values. Algorithms could be created in a way that these personal moral values can be taken into account. But to do so, a method is needed to make machines understand morality and ethics.

Research Questions

The goal of this research topic is the extraction of moral values in a way that machines can understand morality and can base their decisions on certain moral values. This should then be applied according to a certain user and the users‘ personal moral values. In order to achieve this, Machine Learning Algorithms can be used. The following research questions can be derived from that goal:

1. How can moral values be defined formally and represented logically?
2. What Machine Learning Model can be applied to the learning of moral values?
3. What input sources can be used for training the Machine Learning Model?
4. How can the trained Model be applied to one‘s personal moral values?

Approach



1. Formalization

Firstly, the concept of moral values has to be formalized. One possible approach is the assignment of priorities to certain ethical instances (e.g. economic growth vs. enviromental sustainability). These entities of morality can then be weighted against each other and also be compared and brought into conjunction with each other.

2. Thesaurus

The created Machine Learning Model will result in a Thesaurus.
A Thesaurus is “a type of dictionary in which words with similar meanings are arranged in groups“[1].
This Thesaurus will assign a set of correlating moral values to every prioritised ethical instance. Therefore a net of depending moral values will be created in the training of the model.

3. Model Training

The Thesaurus is trained by analysing religious, philosophical and political texts. Providing different points of view and different instances of moral values in correlation to each other, these texts provide the information for the thesaurus.
Approaches for emotion recognition allready exist [2]. This can be extended to recognize moral implications in text.

4. Application

By applying, some moral values of a user can be determined by letting the user do a test or survey and collecting moral points of view. This information can be interpreted as a prioritised entry in the thesaurus. It is now possible to assess correlating moral values of this user. This information can then be used to provide personalized ethics in the algorithm.

Anjo Seidel
IT-Systems Engineering Master
Practical Data Engineering (WiSe19/20)

Hasso Plattner Institute, Potsdam, Germany

E-Mail: anjo.seidel@student.hpi.de

Sources:
[1] https://dictionary.cambridge.org/dictionary/english/thesaurus
[2] Saima Aman; Stan Szpakowicz: “Using Roget’s Thesaurus for Fine-grained Emotion Recognition”
[3] U.Güntzer; G.Jüttner; G.Seegmüller; F.Sarre: “Automatic thesaurus construction by machine learning from retrieval sessions”

Icons:
https://muhimu.es/diversidad/dilema-tranvia-significado/
http://getdrawings.com/type-icon#type-icon-11.png
https://www.pngfuel.com/free-png/bjlgp/download
https://www.flaticon.com/free-icon/holy-bible_1655
https://www.flaticon.com/free-icon/torah_85228
https://www.flaticon.com/free-icon/law-book_293957
https://www.flaticon.com/free-icon/binary-code_274412
https://www.pngwave.com/png-clip-art-ndfnu/download
https://www.flaticon.com/free-icon/black-male-user-symbol_44948
https://www.onlinewebfonts.com/icon/220788
https://www.flaticon.com/free-icon/libra_167812

Hasso Plattner Institut