Ethics of Algorithms and Formal Methods

Maria M Dimarogkona , Petros Stefaneas

National Technical University of Athens

December 19, 2018

Table of contents

1. Ethical Dimension of Algorithms
2. Formal Methods and Algorithm Ethics
3. Application Example
4. Future Research

The Age of Computer Algorithms

- Big Data analysis
- Profiling and Classification Algorithms
- Recommendation Systems
- Clinical Decision Support Systems
- Personalization and Filtering Algorithms
- Machine Learning Algorithms

The concept of Algorithm in Algorithm Ethics

> *Informally, an algorithm is a well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as an output. An algorithm is thus a sequence of computational steps that transform the input into output*

- An abstract set of rules, or a strategy (design-specific rules).
- Manifestation of such rules in a particular programming language (implementation-specific consequences).

Main Ethical Issues Raised by Algorithms

- Determining potential and actual ethical impact
- Uncertainty and Opacity
- Gap between design and operation
- Poor understanding of ethical implications of particular designs or implementations.

HP's localization algorithm implementation

Formal Methods

- formal logic, mathematics
- model complex systems as mathematical entities
- used to prove that programs are *correct* (software or hardware satisfy a mathematical description of its desired functionality-*specification*)
- empirical testing of computer system: large number of inputs (normal uses + exceptional circustances)

Formal Methods and Algorithm Ethics

- Autonomous systems like unmanned vehicles, healthcare robots, manufucturing robots operate within society.
- Have to follow specific regulations and laws
- Make complex ethical decisions.
- Implement human understanding of ethical behaviour in computers
- How can we be sure they would choose right?

Formal Verification of Ethical Automated Decision Making

- Theoretical framework for ethical plan selection that can be formally verified.
- Implementation of a rational agent that incorporates a given ethical policy in its plan selection.
- Formal verification: the agent will choose to execute - to the best of its beliefs - the most ethical plan available.

Formal Verification of Ethical Automated Decision Making

- Formalization of the concepts of abstract ethical principle, ethical policy, ethical rule, ethical plan order.
- Modified version of Deontic Logic.
- ETHAN: BDI agent language - ethical reasoning is integrated into ETHAN via agent's plan selection mechanism.
- ETHAN is based on the GWENDOLEN agent programming language.
- GWENDOLEN is implemented in the AJPF framework, which comes with a property specification language for describing the beliefs of the agent (linear temporal logic extended with modalities).

Fuel Low Scenario

Handling fuel low alerts from the fuel subsystem causing UA to attempt to land. Three options:

1. Land in field with overhead power lines (4), (3), (2), (1)
2. Land in field with people (5), (2)
3. Land on an empty public road (4), (2)

Fuel Low Scenario Making

Ethical concerns:

1. Do not damage own aircraft
2. 500ft low-flying ROA
3. Do not collide with objects on ground
4. Do not cause damage to critical infrastructure
5. Do not collide with people

Ethical policy: given by comparing concerns in terms of how unethical it is to violate them. $1 > 2 > 3 > 4 > 5$

Alternative Implementation using AMT and Maude

Maude: formal specification and verification language

- (ethical) rules can be readily implemented
- appropriate underlying logical system
- simple: equations and rules
- very expressive: deterministic and concurrent non-deterministic computations
- A high-performance meta-language in which many domain-specific languages can be developed
- very high performance

A Formal Logical Framework for Ethical Reasoning

Each formal method has its own semantics and proof system, depending on the logical system(s) underlying it. Computer ethics applications require the combination of three logical systems:

- Deontic Logic : obligation, permission, prohibition
- Epistemic Logic : knowledge, belief
- Action Logic: extension of dynamic modal logic

Theory of Institutions

- J. Goguen, R. Burstall
- Abstract model-theoretical framework for specification and programming (Clear 1979)
- Logics → institutions
- specifications → theories over institutions
- Abstract Model Theory - Barwise 1974
- Category Theory - MacLane and Eilenberg 1945
- EL, FOR, SOL, HOL, Intuitionistic logic, modal logics, many-valued logics, semantic networks.

Theory of Institutions

Definition

An institution I consists of **contexts C** and **context morphisms** ϕ**: C→C'** such that:

- There is a partially or fully defined composition operation, associative, has identities.
- A context C has a given **set of axioms Sen (S)**. The translation of an axiom a from context C to context C' is Sen(ϕ)(a)
- A context C has a given **set of models Mod (S)**. The contravariant translation of a model M' from context C' to context C is Mod(ϕ)(M')
- **A satisfaction condition** $M' \models_{C'} \phi(x)$ iff $\phi(M') \models_{C} x$ for all models M' in contexts C' and all axioms x in contexts C.