Universidade do Minho
Escola de Engenharia

Semana da Escola de Engenharia
October 24 - 27, 2011

A STEP TOWARDS MEDICAL ETHICS MODELLING

Gabriel Pontes
Universidade do Minho, CCTC, Departamento de Informática, Braga, Portugal
E-mail: gabrielpontes@amave.pt

ABSTRACT

Over viewing the evolution of technology and information systems thematic, a trend of growing pro-activeness and limited intelligence is pushing the role of virtual entities, on a step-by-step basis, higher and higher. Many activities are nowadays performed by automated entities, while supervised by humans. Although most of these virtual entities are still rather limited in learning, adaptation and autonomy, displaying solely reactance to predicted or programmed events, several threads of Artificial Intelligence (AI) research methodologies for imbedding further intelligence.

The notion of virtual entity is here used to differentiate entities with higher levels of autonomy, learning, prediction and decision from a mainly reactive and controlled machine. Moreover, considering developments in the area of informatics and AI in particular, it must be considered that many of these entities can exist within a single physical machine or even that a single one can be distributed within limitless machines. Therefore, the notion of a virtual entity in this case is similar to the concept of an agent in the area of Multi-Agent Systems.

As virtual entities become more complex and hold critical functions, a justified doubt and concern regarding the impact of actions performed by these entities arises. From the numerous scenarios where they can interact with their surrounding environment, some carry moral consequences and describe ethically intricate actions from a human point of view. From the need to prevent immoral decisions and ensure confidence regarding these virtual entities, further understanding of the capacity of moral agency, moral modeling and the complexity human moral ethics.

Modeling machine ethics can result in further understanding of human ethics itself, either by defining rules and exceptions, or by knowledge extraction, case classification and patterns search over existing cases and outcomes using different algorithms. One can in fact consider that from the numerous methodologies that exist for the study of moral capacity, for each of them different subsequent potential outcomes can be found. While modeling ethics based on defined moral principles can help defining ethical principles and validate the resulting decision process, using learning algorithms and knowledge extraction over existing moral cases and outcomes can deepen the understanding of the underlying moral rules and patterns that may go unnoticed, but define moral decisions. In other words, theses processes aiming to analyze the essence of morality can be used not only to study their simulation/emulation, but also to deepen and evaluate the moral standards and dilemmas in ethically complex systems. The results from these systems are not limited outcome decisions before an ethical complex problem. Using a perspective of decision support or decision optimization, from a knowledge-base (either by previous studied cases or expert input), bearing in mind a specific scenario, similar cases can be aggregated for human user consideration, rules/principles involved in the decision can be induced with a certain degree of certainty, or conditions can be abduced.

There exists no definite solution for modeling ethical virtual entities, and presently several approaches are being presented and some compared against one another. Studying the present study and investigation in the area, different methodologies for modeling moral capabilities using AI techniques can be segmented according to their main characteristics. One of the most definite and important disparity in methodologies is the usage of explicit reasoning versus black-box reasoning. In explicit reasoning, the processes underneath a moral decision are clearly defined as principles, rules, exceptions, or other structure defined for one particular modeling. When analyzing AI techniques derivatives of symbolic, sub-symbolic or statistical approaches, there exist some that are able to represent their “line of

though", allowing a transparent view of the moral decision process.

One of these techniques is logic programming, in which horn clauses contain the formalisms that mold the reasoning within an existing logical predicate. Current research indicates that non-monotic logic, due to its ability to implement defeasible inference, enabling moral principles to add and still diminish the set of conclusions determined by the knowledge base, is an interesting and promising technique to model moral reasoning. By this mean, principles of benevolence and non-malificience can exist in accordance with other principles that are against their value or state an exception for superseding context principles. Regardless of the use of deductive, inductive or abductive logic, the rules used or attained are explicitly defined. However, the usage of each of these techniques of logic programming varies on the objective and context of application.

On the other hand, while using black-box reasonning, the reasoning behind the moral decision itself cannot be perceived in a clear manner. In other words, within the process of a black-box technique, facing a set of inputs, only a set of outputs can be obtained, not the process or reason behind it. That is the case of neural networks, regardless of the methodologies used to attempt to understand the reasoning behind them, the fact remains that no certainty of the processing underneath the trained neural network exists. Although interesting results can be achieved using neural networks trained on existing moral cases and consequently implementing case based reasoning, the understanding of the moral principles within these black boxes is unknown. Different techniques can be used to reverse-engineer neural network's inner structure and imbedded rules, however, the result is not exactly the rules used but rather an induced or a probabilistic set of them. In the end of this reverse-engineer process, it is attained an induced set of rules of a systems that already uses induction or probabilistic methods to train its processsing, revealing a certainty of doubt over the extracted rules.

Another divergence in ethical modeling is the learning process of rules or reasoning methodologies in ethical dilemmas. When considering a specific area such as medicine, most of the existing knowledge essential to model moral reasoning is contained in deontological principles or case studies. In either of these cases the core of this knowledge is based on individuals or panels of experts. In light of these sources, the moral decision model can be developed from existing principles, from learnt principles or from hybridization of both sources. While one can consider existing deontological principles as existing principles, learnt principles are those extracted from existing cases. These machine-learning behaviors applied to ethics are a rather complex theme as principle learning may result in immoral principles and depending of the methodology used it may not be possible to clearly understand the underlying principles (e.g. black-box machine learning). Inductive logic programming has also expressed in existing research potential to induce principles and their relations from experts reasoning.

When modeling moral behavior in virtual entities, researchers must always bare in mind the environment that molds its principles. For research purposes selection of an area and a purpose is of the essence in order to evaluate results and contextualize the used approach. With this in mind, the disparity between ideal and real environments in the medical arena creates a complex set of scenarios, which are pressing and interesting to analyze from an ethical point of view. Therefore, this project will address moral reasoning in medicine, and apply it in clinical context.

Modeling of ethical reasoning has been a matter of discussion and research among distinct scientific fields, however no definite model has demonstrated undeniable global superiority over the others. However, the context of application of moral reasoning can require one methodology over the other. In areas such as medicine where quality of life and the life itself of a patient may be at stake, the ability to make the reasoning process understandable to staff and to change is of a paramount importance. In this paper we present some of the modeling lines of ethical reasoning applied to medicine, and defend that continuous logic programming presents potential for the development of trustworthy morally aware decision support systems. It is also presented a model of moral decision in two situations that emerge recurrently at the Intensive Care Units, a service where the moral complexity of regular decisions is a motivation for the analyze and development of moral decision support methodologies.

GABRIEL PONTES is a PhD student in Biomedical Engineering (Medical Informatics), University of Minho, Portugal, working in modelling medical ethics.