

Ethics in norm decision making

Marc Serramia

Artificial Intelligence Research Institute (IIIA-CSIC)
Campus UAB, Bellaterra, Barcelona
marcserr@iiia.csic.es

Abstract

Norms are an instrument to coordinate societies, but deciding which norms to enact is a difficult task. Not only norms might have incompatibilities between themselves, such as norms contradicting other norms, but also the cost of implementation can be an important aspect to consider. Furthermore, due to the growing social interest in ethics and the ethical impact norms can have, this ethical implications should also be examined during the decision making process.

Introduction: Deciding norms systems

In (López-Sánchez et al. 2017) we create a method to select norms in an intelligent manner and taking into account some desirable properties. We therefore introduce the concept of norm net, which we define as a pair $NN = \langle N, R \rangle$ of a set of norms N and a set of relations between those norms R . The relations that we consider are exclusivity relations (when two norms are mutually exclusive, they cannot be enacted together); generalization relations (when a norm is more general than another enacting both of them will lead to redundancy); and substitutability relations (when two norms can be substituted). We state that the selected norm system $\Omega \subset NN$ should be sound, that is to not have mutually exclusive norms and neither have redundancy.

We want to maximise the selection by considering as many norms as possible, but since we want to avoid redundancy, the process should aim at selecting the most general norms. In order to discern which norms are more general than others we use the representation power function. This function $r : N \rightarrow \mathbb{R}$ applies to norms and returns a value, the higher the number, the more general the norm.

Furthermore, we think that norms in a real world scenario will have a cost to implement. Therefore we also add a cost function $c : N \rightarrow \mathbb{R}$, and a budget b to stick to.

Having set all these basic definitions we define our problem as finding a sound norm system $\Omega \subset NN$ with maximum representation power and minimum implementation cost, which should be under the given budget b .

In order to solve this problem, we code it as a linear problem that can be solved with state of the art solvers (CPLEX, Gurobi...)

Adding moral values to the decision

Due to the current social interest in ethics and the possible ethical implications norms might have, we found necessary to expand our decision making process taking into account this trait. In (López-Sánchez et al. 2018), we introduce moral values, which we define as abstract objects that help us discern right from wrong. We say that norms are related to moral values, therefore norms will promote the values which they are related to.

To decide which norms are preferred we require the decision maker to input a set of moral values V and a related total ordering (with no ties) of the values $v_1 \succ \dots \succ v_{|V|}$. Hence, in this case v_1 is the most preferred value, while $v_{|V|}$ is the least preferred one. From this given order we build a utility function $u : V \rightarrow \mathbb{R}$ defined as $u(v_i) = 1 + \sum_{k=i+1}^{|V|} u(v_k)$, that weighs the desirability of the values: the higher this utility function the better the value will be.

Since norms might support different values we will find the overall value utility of a norm n as the sum of the utilities of the values it supports: $u(n) = \sum_{v \in \text{val}(n)} u(v)$, where $\text{val}(n)$ stands for the values related to n .

Having defined this new notion we can add it to the previously defined problem. Therefore the problem now is to find a sound norm system $\Omega \subset NN$ with maximum representation power and minimum cost, whose cost is under the given budget b with the largest value utility possible.

Current work: Improving the decision process

We are currently enriching this ethical norm decision problem by means of considering the relations between norms and values as some degree of promotion/demotion, this idea is more accurate as norms can support values in different degrees and can also demote values in different degrees. For example, permitting to build a building on park demotes the value of ecology, but it does not demote it as much as permitting to cut a large forest. We have formalised this concept by assigning a weight in $[-1, 1]$ to norm-value relations, being 1 total promotion and -1 total demotion.

Another improvement we introduce is the use of partial ordering. We think this is an important improvement since not only enables us to define equally preferred values, but also enables us to give partial orders in which values might

not be related to all other values (which could translate to incomparable values).

We are also testing our model to assess its performance and find which factors make the optimisation problems harder.

Future work

Although we have researched deeply to build our current ethical norm decision process and we even have a functional implementation, there is room for improvement.

On the theoretical side, we plan to propose a formal definition of moral value. Although we understand that it is an abstract concept, we would like to find a way to define it more precisely. Furthermore, we have assumed that values are just objects to which some norms might relate to (either by promoting the value or by demoting it), but we have not considered the possibility of values being related to one another. If norms can be mutually exclusive, values might also have this kind of relations. Even more, these possible relations between values can be a cause of the relations between norms.

Another important issue is to find ways to elicitate preferences on values. We think these preferences might depend on the scenario as well as on the society in which the norms are being applied. A good way to obtain these preferences from the society might be to "mine" them from the people's input. How to do so and how to later combine all input into a general value preference remain as open problems.

On the practical side, we think that our current ethical norm decision implementation is very powerful and can have other applications on other decision making scenarios. We would like to find new applications and study how to adapt the ethical norm decision problem to them.

References

- [López-Sánchez et al. 2017] López-Sánchez, M.; Serramia, M.; Rodríguez-Aguilar, J. A.; Morales, J.; and Wooldridge, M. 2017. Automating decision making to help establish norm-based regulations. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017*, 1613–1615.
- [López-Sánchez et al. 2018] López-Sánchez, M.; Serramia, M.; Rodríguez-Aguilar, J. A.; Morales, J.; Wooldridge, M.; and Ansotegui, C. 2018. Exploiting moral values to choose the right norms. In *Proceedings of the 1st Conference on Artificial intelligence, ethics and society, AIES 2018, New Orleans, USA, February 2-3, 2018*, to appear.