Learning Context-Sensitive Norms under Uncertainty

Vasanth Sarathy
vasanth.sarathy@tufts.edu
Tufts University
Medford, Massachusetts

CCS CONCEPTS

• **Computing methodologies** → **Knowledge representation and reasoning**; **Uncertainty quantification**; *Online learning settings*.

KEYWORDS

norm learning; Dempster-Shafer theory; agent-based simulation; uncertainty

ACM Reference Format:
Vasanth Sarathy. 2019. Learning Context-Sensitive Norms under Uncertainty. In *2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES'19), January 27–28, 2019, Honolulu, HI, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3306618.3314315

1 INTRODUCTION

Norms and conventions play a central role in maintaining social order in multi-agent societies [2, 5]. I study the problem of how these norms and conventions can be learned from observation of heterogeneous sources, under conditions of uncertainty. This is necessary as it is not enough to simply hard code a set of norms into a new agent prior to entering society because norms can evolve over time as agents enter and leave the society [9].

What makes learning norms from observation especially challenging are that the same behavior that comprises the norm might have different normative statuses in different situations. For example, yelling out loud is allowed on a beach but not so in a library. Moreover, observing behaviors and gleaning normative status from them is also difficult because the learner is susceptible to different forms of uncertainty. For one, the source (i.e., the agent being observed) may be violating norms. Besides this inherent uncertainty, the learner must also be able to handle its own epistemic uncertainty arising from unreliable sensors or ambiguous observations. The learner must also be able to take into account sanctioning of violations and adjust its normative knowledge accordingly. Finally, the learner must be able to express norms explicitly and in a way that disentangles its own norm compliance, allowing it the choice to either comply or not, and thereby maintain its autonomy.

My work addresses the problem of learning context-sensitive norms in uncertain learning settings. In particular, I represent norms explicitly, and employ a Dempster-Shafer theoretic uncertainty processing framework.

ACM ISBN 978-1-4503-6324-2/19/01.
https://doi.org/10.1145/3306618.3314315

2 RELATED WORK

Norm learning literature is vast and a nice synthesis can be found in [2, 9]. Contemporary approaches vary widely in their choices of representations from explicit logics [1] to implicit policies or reward functions within a state-transition system [3]. As noted earlier, explicit representations have the benefit of allowing agent compliance autonomy. Related works using explicit representations, either do not represent uncertainty in any form [1] or do not account specifically for epistemic uncertainty, or make certain assumptions about the observed agents [4, 6]. Realistically, the learner might not know much about the agent it is observing (i.e., goals or plan libraries). Thus, there are open questions about the role of context and uncertainty when learning norms, that I have begun addressing, as described below.

3 NORM REPRESENTATION AND EPISTEMIC UNCERTAINTY

3.1 Dempster-Shafer Theory

DS-theory is a measure-theoretic mathematical framework that allows for combining pieces of uncertain evidential information to produce degrees of belief for the various events of interest [10]. DS-theory generalizes Bayesian theory, and unlike Bayes, it can represent set-valued random variables, allowing it to capture interval probabilities measures, without committing to particular distributions. Crucially, this allows DS-theory to capture the uncertainty arising not only from randomness in the data, but also from ambiguity and ignorance due to limitations of the agent's sensors and occlusions in the environment.

3.2 Context-Sensitive Norm Representation

We define a context-sensitive behavior $\mathcal{B}$ as:

$$\mathcal{B} \stackrel{\text{def}}{=} C_{\psi}^{[\alpha,\beta]} \tag{1}$$

where C and ψ are ground atomic formulas representing context condition (C), and the corresponding behavior (ψ) possible in that context. The interval $[\alpha, \beta]$ captures the DS-theoretic uncertainty (or probability) that ψ *occurs* in context C, where α represents the belief and β represents the plausibility as defined under DS-theory.

4 RESULTS

4.1 Computational Model of Human Norm Representations

First, with my collaborators, I explored the basic properties of norm processing in humans (representation, activation and learning), and took the first steps towards describing a computational model that could account for these properties. In [7], we showed with experimental data that context-sensitivity of norms is strongly present in humans. We then developed a model – representation format similar to (1) and algorithm – for automatically learning context-sensitive norms from the human data. The model allowed norms to be learned from different types of evidence sources in different contexts, and explicitly captured uncertainty due to variations in the source's reliability and the quality of the evidence. The proposed representation and learning techniques provided a promising platform for studying, computationally, a wide array of cognitive properties of norms.

4.2 Modeling Dynamic Context-Shifting

We then asked the question of how we can account for differences among norm learners and allow distinct agents to communicate and resolve their differences? Learning in the real world is far from perfect; data are often obtained in a streaming, unfolding manner through a series of observations made during a certain time window. Observations are made in context, the identity of which might be uncertain and may even change over time. In [8], we extended our algorithm to track how these norms are learned when observations are made in dynamic situations involving changing contexts and general contextual uncertainty. The revised model allowed us to capture this contextual uncertainty, which in turn, can influence the agent's normative beliefs themselves, resulting in individual differences in normative behavior.

4.3 Extending to Multi-Agent Setting

In our most recent work (under review), we extended the model into an increasingly realistic multi-agent setting, in which we accounted for sanctioning, variable agent compliance, and learning from heterogeneous agents. This work improved on prior algorithms to make it more computationally efficient and tractable for larger sets of norms over longer lifetimes. Moreover, we explored what it means to be a norm and differentiate between conventions and norms as they relate to behaviors. We provided an agent model, series of algorithms, and results from agent-based simulations showing how behavioral regularities can be recognized in a multi-agent setting. The approach showed state of the art performance while overcoming limitations of other approaches. The agent-based modeling approach enabled us to study the impact of four different forms of uncertainty associated with norm compliance rates, sanctioning rates, sensor reliability and environmental occlusion.

5 RESEARCH PLAN AND FUTURE WORK

My research approach is to combine mathematical analysis, algorithm design, computer simulations, robotic implementation, and human-subject experiments to develop and test computational models for normative and ethical behavior. A limitation of the current work is that the learner knows in advance the set of behaviors applicable in a context. In future work, I intend to harness an exciting aspect of DS-theory that allows for expanding and growing the hypothesis space (i.e., set of norms) incrementally and in *real time*. I also intend to run more elaborate agent-based simulations in which I plan to explore (a) the role of sanctions on compliance, (b) how agents that can track norms across multiple contexts, (c) norm compliance and conflicts, and (d) more expressive representations that capture compositional and temporal aspects of norms.

6 CONCLUSION

In this work, I have taken the first steps towards a more complete understanding of the context-sensitivity of norms and how these norms can be learned from observation under conditions of aleatory as well as epistemic uncertainty.

REFERENCES

[1] Giulia Andrighetto, Marco Campenni, Federico Cecconi, and Rosaria Conte. 2010. The Complex Loop of Norm Emergence: A Simulation Model. (2010), 19–35.
[2] Tina Balke, Célia da Costa Pereira, Frank Dignum, Emiliano Lorini, Antonino Rotolo, Wamberto Vasconcelos, and Serena Villata. 2013. Norms in MAS: Definitions and related concepts. In *Dagstuhl Follow-Ups*, Vol. 4. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
[3] Rahmatollah Beheshti, Awrad Mohammed Ali, and Gita Sukthankar. 2015. Cognitive Social Learners: An Architecture for Modeling Normative Behavior. *Twenty-Ninth AAAI Conference on Artificial Intelligence Cognitive* (2015), 2017–2023.
[4] Stephen Cranefield, Felipe Meneguzzi, Nir Oren, and Bastin Tony Roy Savarimuthu. 2016. A Bayesian approach to norm identification. *Frontiers in Artificial Intelligence and Applications* 285 (2016), 622–629.
[5] Bertram F. Malle, Matthias Scheutz, and Joseph L. Austerweil. 2017. Networks of social and moral norms in human and robot agents. In *A World with Robots*, Maria Isabel Aldinhas Ferreira, Joao Silva Sequeira, Mohammad Osman Tokhi, Endre E. Kadar, and Gurvinder Singh Virk (Eds.). Springer, Cham, Switzerland, 3–17.
[6] Nir Oren and Felipe Meneguzzi. 2013. Norm identification through plan recognition. In *Proceedings of the workshop on Coordination, Organization, Institutions and Norms in Agent Systems (COIN 2013@ AAMAS)*.
[7] Vasanth Sarathy, Matthias Scheutz, Joseph Austerweil, Yoed Kenett, Mowafak Allaham, and Bertram Malle. 2017. Mental Representations and Computational Modeling of Context-Specific Human Norm Systems. In *Proceedings of Cognitive Science*.
[8] Vasanth Sarathy, Matthias Scheutz, and Bertram Malle. 2017. Learning Behavioral Norms in Uncertain and Changing Contexts. In *Proceedings of the 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*.
[9] Bastin Tony Roy Savarimuthu and Stephen Cranefield. 2011. Norm Creation, Spreading and Emergence: A Survey of Simulation Models of Norms in Multi-Agent Systems. *Multiagent and Grid Systems* 7, 1 (2011), 21–54. https://doi.org/10.3233/MGS-2011-0167
[10] Glenn Shafer. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.