Modeling Application Scenarios for Responsible Autonomy using Computational Transcendence

Extended Abstract

Jayati Deshmukh
International Institute of Information Technology, Bangalore
Bangalore, India
jayati.deshmukh@iiitb.org

Nikitha Adivi
International Institute of Information Technology, Bangalore
Bangalore, India
nikitha.adivi@iiitb.ac.in

Srinath Srinivasa
International Institute of Information Technology, Bangalore
Bangalore, India
sri@iiitb.ac.in

ABSTRACT

With the prevalence of autonomous agents which should act responsibly, multiple computational models of responsible autonomy have been proposed. We explore the use of one such model called *Computational Transcendence* (CT) which is based on modeling an *elastic sense of self* as a means for emerging responsible autonomy. We discuss how this model can be applied to realistic applications. The first application is on decision-making in multi-agent supply chains, and the second is on adaptive signalling in a road network. In both these applications, we compare CT with several baseline models and find improvement across multiple application-specific metrics. Through this paper, we aim to foster increased research interest in computational transcendence, as a means for architecting responsible multi-agent autonomy for different real-world applications.

KEYWORDS

Responsible AI; Ethics; Autonomy; Multi-agent Systems; Identity

ACM Reference Format:

Jayati Deshmukh, Nikitha Adivi, and Srinath Srinivasa. 2023. Modeling Application Scenarios for Responsible Autonomy using Computational Transcendence: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Across a variety of multi-agent applications, it is becoming crucial to incorporate models of responsibility and ethics in agent design [13, 14]. Autonomous agents which make their own choices should be designed such that they are aware of the consequences of their choices not just on themselves but also on other agents in the system. In closed-world systems with known parameters and constraints, it is possible to create normative systems of responsibility. However, most realistic applications like traffic, supply chains etc. are open-world systems where all the factors cannot be computed in advance.

Machine ethics is a broad area and one of its sub-field is Responsible AI [4, 5, 8, 9, 15]. Different approaches are used to design responsible agents [1, 17]. Top-down approaches enforce responsible behaviour in agents in the form of norms and rules which should be followed by the agents [2]. On the other hand, bottom-up approaches inculcate intrinsic models of responsibility in agents which the agents can learn using reinforcements from the system. There also exist hybrid techniques which combine both these approaches.

Along with theoretical models of responsibility, it is also equally important to design responsible agents operating in specific realistic setups. In this paper, we explore the use of a recently proposed approach called *Computational Transcendence* (CT) [7] as a framework to design intrinsically responsible agents. We present an algorithm to apply Computational Transcendence to any multi-agent application and demonstrate how this framework can be used to design responsible agents in supply chains and traffic management [6].

2 COMPUTATIONAL TRANSCENDENCE

Computational transcendence [7] endows agents with an *elastic sense of self* such that they can *identify* to different extents with other agents, collectives and concepts in the system. Identifying with an external entity results in the agent incorporating the interests of that entity into its own utility function. Formally, the sense of self of an agent a is modeled as follows: $S(a) = (I_a, d_a, \gamma_a)$ Here, I_a represents the identity set which consists of entities (agents, collectives and concepts) which agent a identifies with. d_a represents semantic distance of agent a to every aspect in its identity set. And γ_a represents the level of elasticity or transcendence of the agent. Agent a, with elasticity γ_a identifies with an aspect o at distance $d_a(o)$ with an attenuation factor of $\gamma_a^{d_a(o)}$.

Depending on the elastic identity of an agent, the utility it derives is a combination of its own payoffs as well as scaled payoffs derived by entities in its identity set. Let $\pi_i(o)$ denote the payoff of aspect o, then the utility u of agent a when the system is in state i is computed as follows:

$$u_i(a) = \frac{1}{\sum_{\forall o \in I_a} \gamma_a^{d_a(o)}} \sum_{\forall o \in I_a} \gamma_a^{d_a(o)} \pi_i(o) \tag{1}$$

For every interaction, transcended agents account for the impact (net utility) of their actions on themselves as well as on all the entities in their identity set. They update their semantic distances to each entity they identify with, based on the proportional utility they derive from this association of identity. This helps transcended agents behave more responsibly by choosing an action that leads to collective good even in the presence of individually beneficial actions. Algorithm 1 presents a pseudocode for using CT in any multi-agent application, where agents face the dilemma of responsibility between individually beneficial versus good for the collective choices.

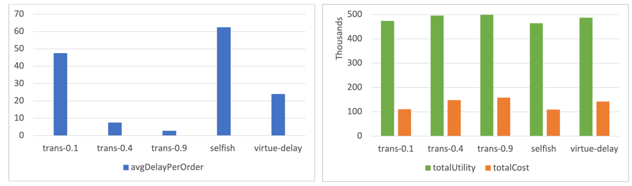
Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

Algorithm 1: Computational Transcendence Pseudocode

Take a system N of n agents
for *agent* $a \in N$ **do**
  Create identity set I_a
  Set transcendence level, $\gamma_a = tl$
  for b *in* I_a **do**
    Set semantic distance $d_a(b) = d$
  end
  $totalpayoff_a = 0$
end
while $TRUE$ **do**
  for *agent* $a \in N$ **do**
    for *action* $k \in actionSet$ **do**
      $u(k) = \frac{\Pi_{a_k}(a) + \sum_{n \in I_a} \gamma_a^{d_a(n)} * \Pi_{a_k}(n)}{1 + \sum_{n \in neig(a)} \gamma_a^{d_a(n)}}$
    end
    $\text{softmax}(actionSet_a) = \frac{\exp(u_a)}{\sum_{k \in actionSet} \exp(u_k)}$
    probabilistically choose action a_T
    $totalpayoff_a += \Pi_{a_T}(a)$
    for *aspect* $b \in I_a$ **do**
      $totalpayoff_b += \Pi_{a_T}(b)$
      $totalpayoff_a += \gamma_a^{d_a(b)} * \Pi_{a_T}(b)$
    end
  end
  `updateDistances()`
end
Function `updateDistances()`:
  for *agent* $a \in N$ **do**
    for *aspect* $b \in I_a$ **do**
      $\Delta d_a(b) = \frac{totalpayoff_a(b)}{totalpayoff_a}$
      $d_a(b) = d_a(b) - lr * \Delta d_a(b)$
    end
  end
  return
End Function

3 CT IN SUPPLY CHAIN MANAGEMENT

A supply chain can be modeled as a multi-agent network of agents which pass orders from source to destination via intermediaries [3, 10, 18]. We have modeled a 1−tier supply chain with a single intermediary for each order. An order is of the form $< s, i, d, ts >$ where s denotes the seller who ships order created at time-step ts to buyer d via intermediary i. The dilemma of an intermediary agent is to decide whether to wait for more orders or to forward existing orders. If it forwards the current batch of orders, it incurs a high shipping cost while the source agents get a utility depending on the delay of their orders. If it waits to accumulate more orders, it incurs a small book-keeping cost of storing all the orders while source agents incur costs since their orders are delayed. Thus, while there is no rational incentive for an intermediary to forward the orders, the system works only if orders are forwarded intermittently.



(a) Average Delay per Order **(b) Total Utility and Cost**

Figure 1: Supply chain metrics of different intermediaries

Intermediate agents can use different strategies for decision making. We have simulated three types of agents: *Selfish agent* forwards only when its book-keeping cost is more than its shipping cost. *Virtue agent - Delay* acts virtuously by minimizing the average delay per order of all its existing orders. *Transcended agent* decides based on its own cost, utility and impact of its action on the source agent, as per Algorithm 1. Figure 1 presents the metrics like average delay per order, total utility and cost for different types of agents. We observe that as transcendence level γ increases, though the total cost increases, total utility also increases and average delay per order reduces leading to higher customer satisfaction. Varying γ generates a broad variety of responsible behaviour as compared to selfish and virtuous agents. Also, γ can be adjusted based on the priorities and requirements of the supply chain system.

4 CT IN TRAFFIC MANAGEMENT

Managing traffic in a network is more of an adaptation instead of an optimization problem [11]. Traffic flows in a network can be made smoother and efficient using adaptive traffic lights (ATLs) [12, 16]. These ATLs can be modeled as autonomous agents which optimize flows. Depending on dynamic traffic flows, ATLs choose to turn green for one of the incoming lanes at an intersection using a variety of ways [19]. We modeled three types of ATLs as follows: *Maximum Waiting Time TL* selects a lane on which vehicles have the maximum waiting time, *Maximum Queue Length TL* selects a lane which has the maximum queue length and *Transcended TL* factors traffic being sent from all its neighbouring intersections and periodically updates its distance to each of its neighbours depending on the change in traffic flow. In this application, we measure metrics like average waiting time and average speed of all the vehicles in the network as a whole. Simulation results on a 4−intersection grid over multiple runs, show promising results where the grid with transcended TLs has lowest waiting time and highest average speed as compared to the grid with other types of ATLs.

5 CONCLUSIONS

Computational transcendence framework is useful to design intrinsically responsible agents across a variety of applications. Also, varying transcendence levels and semantic distances gives the ability to agents to adapt to dynamically changing scenarios. Transcended agents are able to optimize flows (of orders, traffic etc.) at the network level rather than individual agent level. We hope that this work motivates and builds a case for designing responsible autonomous agents using computational transcendence across a diverse set of real-world applications.

ACKNOWLEDGMENTS

Thanks to Machine Intelligence and Robotics (MINRO) center funded by Government of Karnataka, India and Centre for Internet of Ethical Things (CIET) funded by Government of Karnataka, India and World Economic Forum for supporting this work.

REFERENCES

[1] Colin Allen, Iva Smit, and Wendell Wallach. 2005. Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and information technology* 7, 3 (2005), 149–155.

[2] Paul Bremner, Louise A Dennis, Michael Fisher, and Alan F Winfield. 2019. On proactive, transparent, and verifiable ethical reasoning for robots. *Proc. IEEE* 107, 3 (2019), 541–561.

[3] Gerard P Cachon and Serguei Netessine. 2006. Game theory in supply chain analysis. *Models, methods, and applications for innovative decision making* (2006), 200–233.

[4] Roger Clarke. 2019. Principles and business processes for responsible AI. *Computer Law & Security Review* 35, 4 (2019), 410–422.

[5] Mehdi Dastani and Vahid Yazdanpanah. 2022. Responsibility of AI Systems. *AI & SOCIETY* (2022). https://doi.org/10.1007/s00146-022-01481-4

[6] Jayati Deshmukh, Nikitha Adivi, and Srinath Srinivasa. 2023. *Modeling Application Scenarios for Responsible Autonomy using Computational Transcendence.* Technical Report. International Institute of Information Technology, Bangalore.

[7] Jayati Deshmukh and Srinath Srinivasa. 2022. Computational Transcendence: Responsibility and agency. *Frontiers in Robotics and AI* 9 (2022). https://doi.org/10.3389/frobt.2022.977303

[8] Virginia Dignum. 2017. Responsible Autonomy. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17.* 4698–4704. https://doi.org/10.24963/ijcai.2017/655

[9] Virginia Dignum. 2020. Responsibility and artificial intelligence. *The Oxford Handbook of Ethics of AI* 4698 (2020), 215.

[10] D. Frey, P.-O. Woelk, T. Stockheim, and R. Zimmermann. 2003. Integrated multi-agent-based supply chain management. In *WET ICE 2003. Proceedings. Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003.* 24–29. https://doi.org/10.1109/ENABL.2003.1231377

[11] Carlos Gershenson. 2004. Self-organizing traffic lights. *arXiv preprint nlin/0411066* (2004).

[12] Dave McKenney and Tony White. 2013. Distributed and adaptive traffic signal control within a realistic traffic simulation. *Engineering Applications of Artificial Intelligence* 26, 1 (2013), 574–583.

[13] Michal Pěchouček and Vladimír Mařík. 2008. Industrial deployment of multi-agent technologies: review and selected case studies. *Autonomous agents and multi-agent systems* 17, 3 (2008), 397–431.

[14] Michal Pechoucek, Simon G Thompson, and Holger Voos. 2008. *Defence Industry Applications of Autonomous Agents and Multi-Agent Systems.* Springer.

[15] Dorian Peters, Karina Vold, Diana Robinson, and Rafael A Calvo. 2020. Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society* 1, 1 (2020), 34–47.

[16] Arthur G Sims. 1979. The Sydney coordinated adaptive traffic system. In *Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems, 1979, Pacific Grove, California, USA.*

[17] Suzanne Tolmeijer, Markus Kneer, Cristina Sarasua, Markus Christen, and Abraham Bernstein. 2020. Implementations in machine ethics: A survey. *ACM Computing Surveys (CSUR)* 53, 6 (2020), 1–38.

[18] Stefan Voß and David L Woodruff. 2006. *Introduction to computational optimization models for production planning in a supply chain.* Vol. 240. Springer Science & Business Media.

[19] MA Wiering, J van Veenen, Jilles Vreeken, Arne Koopman, et al. 2004. Intelligent traffic light control. (2004).