'Why didn't you allocate this task to them?' Negotiation-Aware Task Allocation and Contrastive Explanation Generation

Extended Abstract

Zahra Zahedi
Arizona State University
zzahedi@asu.edu

Sailik Sengupta*
AWS AI Labs
sailiks@amazon.com

Subbarao Kambhampati
Arizona State University
rao@asu.edu

ABSTRACT

In this work, we design an Artificially Intelligent Task Allocator (AITA) that proposes a task allocation for multi-agent systems especially with humans. A key property of this allocation is that when an agent with imperfect knowledge (about their teammate's costs and/or the team's performance metric) questions the allocation by contesting with a counterfactual, a contrastive explanation is provided to answer their challenge. For this, we consider a negotiation process that produces a negotiation-aware task allocation and, in turn, leverages a negotiation tree to provide a contrastive explanation. With human subject studies, we show that the proposed allocation indeed appears fair to a majority of participants, and the explanations generated are easy to comprehend and convincing.

KEYWORDS

Task Allocation; Contrastive Explanation; Negotiation

ACM Reference Format:

Zahra Zahedi, Sailik Sengupta, and Subbarao Kambhampati. 2023. *'Why didn't you allocate this task to them?'* Negotiation-Aware Task Allocation and Contrastive Explanation Generation: 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

Task allocation is crucial for the smooth functioning of human teams; it involves assigning individual team members to tasks while optimizing a set of predefined metrics, such as skill-sets, capacity, and timing constraints [24]. An array of distributed approaches exist that comprise of online methods which consider negotiation, bargaining and local commitments to arrive at a Pareto optimal allocation [3, 5, 6, 8–11, 19, 21]. While these distributed negotiation-based allocations can make individual agents happy, they tend to compromise on the efficiency and optimality of the solution. In this regard, centralized approaches produce more efficient and optimal allocations [7, 12], but lead to disgruntled agents who may wish to contest the proposed solution. Therefore, a centralized solution needs to foresee this challenge when coming up with a solution and, in turn, be able to provide explanations [2, 16]. Further, [13] highlights that providing explanations is both important and challenging in a multi-agent environment (such as task allocation). While generating explicable solutions [14, 17, 26] and providing explanations [4, 15, 20, 22, 23] has been extensively studied for machine learning and automated planning, these considerations have not been tantamount to optimality and efficiency in multi-agent task-allocation scenarios. To bridge this gap, we propose AITA, an **A**rtificially **I**ntelligent **T**ask **A**llocator that initially leverages a centralized allocation algorithm modeled after negotiation to come up with an allocation that accounts for the costs of individual agents and overall team performance. AITA can then provide contrastive explanations when a proposed allocation is contested using a counterfactual. The explanations have two desirable properties: they have a graphical form that effectively distills relevant pieces of information, and they also act as a certificate which guarantees explicability to the human. We conduct human subject studies in three different task-allocation scenarios and show that the allocations proposed by AITA are deemed fair by the majority of subjects. When users question AITA's allocations and are presented different explanations, they find our explanation to be the most understandable and the most convincing.

*Work done while at Arizona State University.
The full paper can be found on Arxiv [25].

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.

2 PROBLEM FORMULATION

Our task allocation problem can be defined using a 3-tuple $\langle A, T, C\rangle$ where $A = \{0, 1, \ldots, n\}$ where n denotes AITA and $0, \ldots n-1$ denotes the n humans, $T = \{T_1, \ldots, T_m\}$ denotes m indivisible tasks that need to be allocated to the n humans, and $C = \{C_0, C_1, \ldots C_n\}$ denotes the cost functions of each agent. C_n represents the overall performance cost metric associated with a task allocation outcome o. For a task t, we denote the human i's cost for that task as $C_i(t)$. Let O denote the set of allocations and an allocation $o (\in O)$ represent a one-to-many function from the set of humans to tasks . An outcome o can be written as $\langle o_1, o_2, \ldots o_m\rangle$ where each $o_i \in \{0, \ldots, n-1\}$ denotes the human performing task i. Further, let us denote the set of tasks allocated to a human i, given allocation o, as $T_i = \{j : o_j = i\}$. For any allocation $o \in O$, there are two types of costs for AITA: (1) Cost for each human i to adhere to o. $C_i(o) = \Sigma_{j \in T_i} C_i(j)$, and (2) An overall performance cost $C_n(o)$.

Negotiation Tree The negotiation between agents can be represented as a tree where each node (i, o) represents an agent i proposing an outcome o, and other agents can accept or reject it. If rejected, the next agent $i+1$ proposes an outcome that (1) not an offer previously seen in the tree, and (2) is optimal regarding to agent $i + 1$'s cost among the remaining offers. The tree progresses either until all agents *accept* the offer or all outcomes are exhausted. Each negotiation step increases the time needed to reach a final task-allocation. So, similar to [1], we consider a discount factor. The acceptance or rejection of an allocation for agent i is based on whether the cost of the allocation is less than or equal to their cost of a negotiation-aware explicable allocation $O^i_{na\_exp}$.

$$\begin{cases} accept\ o & if\ C_i(o) \leq C_i(O^i_{na\_exp}) \\ reject\ o & otherwise \end{cases}$$

Negotiation-Aware Explicable Allocation

An allocation is considered explicable by all agents iff, upon negotiation, all the agents are willing to accept it. Formally, an acceptable allocation at step s of the negotiation, denoted as $O_{na\_exp}(s)$, has the following properties:

1. All agents believe that allocations at a later step of the negotiation will result in a higher cost for them.

$$\forall i,\ \forall s' > s \qquad C_i(O(s')) > C_i(O_{na\_exp}(s))$$

2. All allocations offered by agent i at step s'' before s, denoted as $O^i_{opt}(s'')$, is rejected at least by one other agent. The *opt* in the subscript indicates that the allocation $O^i_{opt}(s'')$ at step s'' has the optimal cost for agent i at step s''. Formally,

$$\forall s'' < s,\ \exists j \neq i, \quad C_j(O^i_{opt}(s'')) > C_j(O_{na\_exp}(s))$$

We now describe how AITA finds a negotiation-aware explicable allocation. The negotiation process to find an explicable allocation can be viewed as an sequential bargaining game. With the correct information about the costs, and considering the limited computational capability of the humans, AITA uses the simulated negotiation by enumerating all the periods of the sequential bargaining game to find the negotiation-aware explicable allocation that is accepted by all. The optimal solution of our defined bargaining game (negotiation-aware explicable allocation) is the Subgame Perfect Equilibrium of the game that can be found using backward induction [18].

Counterfactual Allocation and Explanation

In situations where a human has complete information and computational capabilities, they will understand that AITA's allocation is explicable and does not need an explanation. However, in real-world settings, humans may not have complete information about other humans' utility functions [21], so a human may perceive AITA's allocation as suboptimal. This can result in the human contesting thier allocation by proposing a counterfactual allocation that they believe would be better for them and accepted by all other players. AITA can generate an effective explanation that refutes the counterfactual allocation by describing a negotiation tree using actual costs. An explanation is a negotiation tree that shows the counterfactual allocation will result in a final allocation with a higher cost for the human than AITA's proposed allocation.

3 EXPERIMENTAL RESULTS

This section examines if the proposed explicable allocation is perceived as fair and if the contrastive explanations generated by AITA are effective. Two experiments are conducted:

Table 1: Users who felt AITA's allocation was fair (% Fair) and the average Understandable (U) and Convincing (C) scores for the various explanations.

Domain	% Fair	Vacuous		Verbose		Neg-tree	
		U	C	U	C	U	C
Cooking	84.2%	4.5	2.33	4.3	4	4.5	4
Class Project	86.4%	4.4	2.8	4.2	3.4	3.8	4.4
Paper Writing	55.0%	-	-	-	-	-	-

Relative Case We presented two task allocation scenarios to 38 participants - (1) cooking at a restaurant with two teammates and three tasks, and (2) dividing five tasks between a senior and junior grad/undergrad student. we presented the participants with AITA's proposed allocation and counterfactual allocations. When the human selects a counterfactual, implying that AITA's proposed allocation is inexplicable, we present them with three explanations: our negotiation-tree based explanation, a *vacuous* explanation that simply states that the human's counterfactual won't be accepted by others and doesn't ensure a good overall performance metric, a *verbose* explanation that provides the cost of all their teammates and the performance metric for all allocations.

Absolute Case Setup In this study, we had 40 human subjects participate in a task allocation scenario where a senior researcher and a junior researcher are working on a paper with three different tasks. Similar to the previous case, the subjects have to select whether the AITA's proposed allocation is fair or select between either of the two counterfactual allocations. In contrast to the previous case, upon selecting a counterfactual, the subject is only presented with the negotiation-tree explanation.

Results. Across the three different domains, a majority of the participants selected that AITA's allocation is fair (see table 1). This shows that our formally defined negotiation-aware explicable allocation does indeed appear fair to humans. For participants who asked for explanations by providing a counterfactual, they were asked to rate the comprehensibility and convincing power of the provided explanations on a scale of $1 - 5$. We observed that the negotiation tree was judged to be *understandable* and *moderately convincing* on average in the absolute setting. In the cumulative setting, results in both the cooking and the class project domain show that our explanation is the most convincing one. It is also perceived as the most understandable explanation (but shared the stage with the Vacuous explanation.)

4 CONCLUSION

In this paper, we proposed a centralized AI task allocator that provides negotiation-aware explicable allocation to human teams using a simulated negotiation-based approach. AITA also gives counterfactual explanations to team members unsatisfied with the proposed allocation. The paper includes human subject studies to show that the allocation is fair to the majority of humans, and the provided explanations are understandable and convincing.

ACKNOWLEDGMENTS

This research is supported in part by ONR grants N00014-16-1-2892, N00014-18-1- 2442, N00014-18-1-2840, N00014-9-1-2119, AFOSR grant FA9550-18-1-0067, DARPA SAIL-ON grant W911NF-19- 2-0006, and a JP Morgan AI Faculty Research grant.

REFERENCES

[1] Sandeep Baliga and Roberto Serrano. 1995. Multilateral bargaining with imperfect information. *Journal of Economic Theory* 67, 2 (1995), 578–589.

[2] Graham L Bradley and Beverley A Sparks. 2009. Dealing with service failures: The use of explanations. *Journal of Travel & Tourism Marketing* 26, 2 (2009), 129–143.

[3] Steven J Brams and Alan D Taylor. 1996. *Fair Division: From cake-cutting to dispute resolution.* Cambridge University Press.

[4] Tathagata Chakraborti, Sarath Sreedharan, and Subbarao Kambhampati. 2020. The Emerging Landscape of Explainable AI Planning and Decision Making. *arXiv preprint arXiv:2002.11697* (2020).

[5] Yann Chevaleyre, Paul E Dunne, Ulle Endriss, Jérôme Lang, Michel Lemaitre, Nicolas Maudet, Julian Padget, Steve Phelps, Juan A Rodriguez-Aguilar, and Paulo Sousa. 2006. Issues in multiagent resource allocation. *Informatica* 30, 1 (2006).

[6] Yann Chevaleyre, Ulle Endriss, and Nicolas Maudet. 2010. Simple negotiation schemes for agents with simple preferences: Sufficiency, necessity and maximality. *Autonomous Agents and Multi-Agent Systems* 20, 2 (2010), 234–259.

[7] P Cramton. 2006. Introduction to combinatorial auctions. P. Cramton, Y. Shoham, R. Steinberg, eds., Combinatorial Auctions.

[8] Ulrich Endriss, Nicolas Maudet, Fariba Sadri, and Francesca Toni. 2003. On optimal outcomes of negotiations over resources. In *AAMAS*, Vol. 3. 177–184.

[9] Ulrich Endriss, Nicolas Maudet, Fariba Sadri, and Francesca Toni. 2006. Negotiating socially optimal allocations of resources. *Journal of artificial intelligence research* (2006).

[10] Sefi Erlich, Noam Hazon, and Sarit Kraus. 2018. Negotiation strategies for agents with ordinal preferences. *arXiv preprint arXiv:1805.00913* (2018).

[11] Shaheen Fatima, Sarit Kraus, and Michael Wooldridge. 2014. The negotiation game. *IEEE Intelligent Systems* 29, 5 (2014), 57–61.

[12] Luke Hunsberger and Barbara J Grosz. 2000. A combinatorial auction for collaborative planning. In *Proceedings fourth international conference on multiagent systems*. IEEE, 151–158.

[13] Sarit Kraus, Amos Azaria, Jelena Fiosina, Maike Greve, Noam Hazon, Lutz Kolbe, Tim-Benjamin Lembcke, Jörg P Müller, Sören Schleibaum, and Mark Vollrath. 2019. AI for Explaining Decisions in Multi-Agent Environments. *arXiv preprint arXiv:1910.04404* (2019).

[14] Anagha Kulkarni, Yantian Zha, Tathagata Chakraborti, Satya Gautam Vadlamudi, Yu Zhang, and Subbarao Kambhampati. 2019. Explicable planning as minimizing distance from expected behavior. In *AAMAS*.

[15] David Alvarez Melis and Tommi Jaakkola. 2018. Towards robust interpretability with self-explaining neural networks. In *Advances in Neural Information Processing Systems*. 7775–7784.

[16] Tim Miller. 2018. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* (2018).

[17] Alberto Olmo, Sailik Sengupta, and Subbarao Kambhampati. 2020. Not all failure modes are created equal: Training deep neural networks for explicable (mis) classification. *arXiv preprint arXiv:2006.14841* (2020).

[18] Martin J Osborne et al. 2004. *An introduction to game theory*. Vol. 3. Oxford university press New York.

[19] Noam Peled, Sarit Kraus, et al. 2013. An agent design for repeated negotiation and information revelation with people. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

[20] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 1135–1144.

[21] Sabyasachi Saha and Sandip Sen. 2007. An Efficient Protocol for Negotiation over Multiple Indivisible Resources.. In *IJCAI*, Vol. 7. 1494–1499.

[22] Sarath Sreedharan, Siddharth Srivastava, and Subbarao Kambhampati. 2018. Hierarchical Expertise Level Modeling for User Specific Contrastive Explanations.. In *IJCAI*. 4829–4836.

[23] Karthik Valmeekam, Sarath Sreedharan, Sailik Sengupta, and Subbarao Kambhampati. 2022. RADAR-X: An Interactive Mixed Initiative Planning Interface Pairing Contrastive Explanations and Revised Plan Suggestions. In *Proceedings of the International Conference on Automated Planning and Scheduling*, Vol. 32. 508–517.

[24] D Michael Warner and Juan Prawda. 1972. A mathematical programming model for scheduling nursing personnel in a hospital. *Management Science* 19, 4-part-1 (1972), 411–422.

[25] Zahra Zahedi, Sailik Sengupta, and Subbarao Kambhampati. 2020. Why didn't you allocate this task to them?'Negotiation-Aware Task Allocation and Contrastive Explanation Generation. *arXiv preprint arXiv:2002.01640* (2020).

[26] Yu Zhang, Sarath Sreedharan, Anagha Kulkarni, Tathagata Chakraborti, Hankz Hankui Zhuo, and Subbarao Kambhampati. 2017. Plan explicability and predictability for robot task planning. In *ICRA*. IEEE, 1313–1320.