AFRL-AFOSR-VA-TR-2023-0156

Implicit Communication in Human-Machine Collaboration

Dragan, Anca
REGENTS OF THE UNIVERSITY OF CALIFORNIA
2150 SHATTUCK AVE RM 313
BERKELEY, CA,
US

11/10/2022
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20221110	**2. REPORT TYPE** Final	**3. DATES COVERED** **START DATE** 20170801	**END DATE** 20220731

4. TITLE AND SUBTITLE
Implicit Communication in Human-Machine Collaboration

5a. CONTRACT NUMBER	**5b. GRANT NUMBER** FA9550-17-1-0308	**5c. PROGRAM ELEMENT NUMBER** 61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Anca Dragan

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) REGENTS OF THE UNIVERSITY OF CALIFORNIA 2150 SHATTUCK AVE RM 313 BERKELEY, CA US	**8. PERFORMING ORGANIZATION REPORT NUMBER**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/AFOSR RTA2	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** AFRL-AFOSR-VA-TR-2023-0156

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of this project was to advance human-machine collaboration, specifically focusing on the machine's ability to understand and partake in implicit communication. We focused much of the work on inferring underlying human objectives and preferences, from learning from different types of input (physical corrections, feature traces, even one single state), to introducing better models of human behavior that lead to better inference, to preventing wrong/harmful inference. We also made a number of contributions on enabling robots to communicate themselves implicitly: objectives, capability/incapability, their future task plans, and even emotional state. We have published numerous conference and journal papers, and received 4 best paper nominations (three times at HRI, the premier conference in human-robot interaction, and once at TRO).

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT** UU	**18. NUMBER OF PAGES** 6
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U		

19a. NAME OF RESPONSIBLE PERSON HAL GREENWALD	**19b. PHONE NUMBER** *(Include area code)* 312-425-8441

DISTRIBUTION A: Distribution approved for public release.

Standard Form 298 (Rev.5/2020)
Prescribed by ANSI Std. Z39.18

FA9550-17-1-0308
Implicit Communication in Human-Machine Collaboration

PI: Anca Dragan

Abstract

The goal of this project was to advance human-machine collaboration, specifically focusing an the machine's ability to understand and partake in implicit communication. We focused much of the work on inferring underlying human objectives and preferences, from learning from different types of input (physical corrections, feature traces, even one single state), to introducing better models of human behavior that lead to better inference, to preventing wrong/harmful inference. We also made a number of contributions on enabling robots to communicate themselves implicitly: objectives, capability/incapability, their future task plans, and even emotional state. We have published numerous conference and journal papers, and received 4 best paper nominations (three times at HRI, the premier conference in human-robot interaction, and once at TRO).

1 Objectives

Our objectives were to enable inference of human internal states beyond goal configurations, and to enable expression from machines to humans beyond goal configurations. Specifically, we targeted preferences, task plans, constraints, capabilities, and emotions.

2 Accomplishments

Inference beyond goal configurations

Inference of objectives and preferences

Understanding the communication implicit in physical interaction [CoRL 2017, oral]
[Bajcsy et al.(2017)Bajcsy, Losey, O'Malley, and Dragan]

When humans and robots work in close proximity, physical interaction is inevitable. Traditionally, robots treat physical interaction as a disturbance, and resume their original behavior after the interaction ends. In contrast, we argue that physical human interaction is informative: it is useful information about how the robot should be doing its task. We formalize learning from such interactions as a dynamical system in which the task objective has parameters that are part of the hidden state, and physical human interactions are observations about these parameters. We derive an online approximation of the robot's optimal policy in this system, and test it in a user study. The results suggest that learning from physical interaction leads to better robot task performance with less human effort.

Perfecting our understanding of physical interaction communication with more realistic human models [HRI 2018]
[Bajcsy et al.(2018)Bajcsy, Losey, O'Malley, and Dragan]

We focus on learning robot objective functions from human guidance: specifically, from physical corrections provided by the person while the robot is acting. Objective functions are typically parametrized in terms of features, which capture aspects of the task that might be important. When the person intervenes to correct the robot's behavior, the robot should update its understanding of which features matter, how much, and in what way. Unfortunately, real users do not provide optimal corrections that isolate exactly what the robot was doing wrong. Thus, when receiving a correction, it is difficult for the robot to determine which features the person meant to correct, and which features were changed unintentionally. In this paper, we propose to improve the efficiency of robot learning during physical interactions by reducing unintended learning. Our approach allows the human-robot team to focus on learning one feature at a time, unlike state-of-the-art techniques that update all features at once. We derive an online method for identifying the single feature which the human is trying to change during physical interaction, and experimentally compare this one-at-a-time approach to the all-at-once baseline in a user study. Our results suggest that users teaching one-at-a-time perform better, especially in tasks that require changing multiple features.

Inferring preferences implicit in the state of the world [ICLR'19, ICLR'21]
[and. Krasheninnikov et al.(2019)and. Krasheninnikov, Alexander, Abbeel, and Dragan,Lindner et al.(2021)Lindner, Shah, Abbeel, and Dragan]

Reinforcement learning (RL) agents optimize only the features specified in a reward function and are indifferent to anything left out inadvertently. This means that we must not only specify what to do, but also the much larger space of what not to do. It is easy to forget these preferences, since these preferences are already satisfied in our environment. This motivates our key insight: when a robot is deployed in an environment that humans act in, the state of the environment is already optimized for what humans want. We can therefore use this implicit preference information from the state to fill in the blanks. We develop an algorithm based on Maximum Causal Entropy IRL and use it to evaluate the idea in a suite of proof-of-concept environments designed to show its properties. We find that information from the initial state can be used to infer both side effects that should be avoided as well as preferences for how the environment should be organized

Introducing new types of human input for inferring preferences [HRI'21, best paper finalist, IJRR'22]
[Bobu et al.(2021)Bobu, Wiggert, Tomlin, and Dragan, Bobu et al.(2022)Bobu, Wiggert, Tomlin, and Dragan]

When a person is not satisfied with how a robot performs a task, they can intervene to correct it. Reward learning methods enable the robot to adapt its reward function online based on such human input, but they rely on handcrafted features. When the correction cannot be explained by these features, recent work in deep Inverse Reinforcement Learning (IRL) suggests that the robot could ask for task demonstrations and recover a reward defined over the raw state space. Our insight is that rather than implicitly learning about the missing feature(s) from demonstrations, the robot should instead ask for data that explicitly teaches it about what it is missing. We introduce a new type of human input in which the person guides the robot from states where the feature being taught is highly expressed to states where it is not. We propose an algorithm for learning the feature from the raw state space and integrating it into the reward function. By focusing the human input on the missing feature, our method decreases sample complexity and improves generalization of the learned reward over the above deep IRL baseline. We show this in experiments with a physical 7DOF robot manipulator, as well as in a user study conducted in a simulated environment.

Solutions when preferences lie along multiple criteria [NeurIPS'20]
[Bhatia et al.(2020)Bhatia, Pananjady, Bartlett, Dragan, and Wainwright]

The literature on ranking from ordinal data is vast, and there are several ways to aggregate overall preferences from pairwise comparisons between objects. In particular, it is well known that any Nash equilibrium of the zero sum game induced by the preference matrix defines a natural solution concept (winning distribution over objects) known as a von Neumann winner. Many real-world problems, however, are inevitably multi-criteria, with different pairwise preferences governing the different criteria. In this work, we generalize the notion of a von Neumann winner to the multi-criteria setting by taking inspiration from Blackwell's approachability. Our framework allows for non-linear aggregation of preferences across criteria, and generalizes the linearization-based approach from multi-objective optimization. From a theoretical standpoint, we show that the Blackwell winner of a multi-criteria problem instance can be computed as the solution to a convex optimization problem. Furthermore, given random samples of pairwise comparisons, we show that a simple plug-in estimator achieves near-optimal minimax sample complexity. Finally, we showcase the practical utility of our framework in a user study on autonomous driving, where we find that the Blackwell winner outperforms the von Neumann winner for the overall preferences.

Introducing better models for inference

Less is more: rethinking probabilistic models of human behavior [HRI 2020, best paper award] [Bobu et al.(2020b)Bobu, Scobee, Satry, and Dragan]

Robots need models of human behavior for both inferring human goals and preferences, and predicting what people will do. A common model is the Boltzmann noisily-rational decision model, which assumes people approximately optimize a reward function and choose trajectories in proportion to their exponentiated reward. While this model has been successful in a variety of robotics domains, its roots lie in econometrics, and in modeling decisions among different discrete options, each with its own utility or reward. In contrast, human trajectories lie in a continuous space, with continuous-valued features that influence the reward function. We propose that it is time to rethink the Boltzmann model, and design it from the ground up to operate over such trajectory spaces. We introduce a model that explicitly accounts for distances between trajectories, rather than only their rewards. Rather than each trajectory affecting the decision independently, similar trajectories now affect the decision together. We start by showing that our model better explains human behavior in a user study. We then analyze the implications this has for robot inference, first in toy environments where we have ground truth and find more accurate inference, and finally for a 7DOF robot arm learning from user demonstrations.

The assistive multi-armed bandit [HRI 2019] [Chan et al.(2019)Chan, Hadfield-Menell, Srinivasa, and Dragan]

Learning preferences implicit in the choices humans make is a well studied problem in both economics and computer science. However, most work makes the assumption that humans are acting (noisily) optimally with respect to their preferences. Such approaches can fail when people are themselves learning about what they want. In this work, we introduce the assistive multi-armed bandit, where a robot assists a human playing a bandit task to maximize cumulative reward. In this problem, the human does not know the reward function but can learn it through the rewards received from arm pulls; the robot only observes which arms the human pulls but not the reward associated with each pull. We offer sufficient and necessary conditions for successfully assisting the human in this framework. Surprisingly, better human performance in isolation does not necessarily lead to better performance when assisted by the robot: a human policy can do better by effectively communicating its observed rewards to the robot. We conduct proof-of-concept experiments that support these results. We see this work as contributing towards a theory behind algorithms for human-robot interaction.

Where Do You Think You're Going?: Inferring Beliefs about Dynamics from Behavior [NeurIPS 2018] [Reddy et al.(2018)Reddy, Dragan, and Levine]

Inferring intent from observed behavior has been studied extensively within the frameworks of Bayesian inverse planning and inverse reinforcement learning. These methods infer a goal or reward function that best explains the actions of the observed agent, typically a human demonstrator. Another agent can use this inferred intent to predict, imitate, or assist the human user. However, a central assumption in inverse reinforcement learning is that the demonstrator is close to optimal. While models of suboptimal behavior exist, they typically assume that suboptimal actions are the result of some type of random noise or a known cognitive bias, like temporal inconsistency. In this paper, we take an alternative approach, and model suboptimal behavior as the result of internal model misspecification: the reason that user actions might deviate from near-optimal actions is that the user has an incorrect set of beliefs about the rules – the dynamics – governing how actions affect the environment. Our insight is that while demonstrated actions may be suboptimal in the real world, they may actually be near-optimal with respect to the user's internal model of the dynamics. By estimating these internal beliefs from observed behavior, we arrive at a new method for inferring intent. We demonstrate in simulation and in a user study with 12 participants that this approach enables us to more accurately model human intent, and can be used in a variety of applications, including offering assistance in a shared autonomy framework and inferring human preferences.

Preventing wrong/harmful inference

Learning under misspecified objective spaces [CoRL 2018, invited to special issue and TRO 2020, best paper honorable mention]

[Bobu et al.(2018)Bobu, Bajcsy, Fisac, and A.D.Dragan]
[Bobu et al.(2020a)Bobu, Bajcsy, Fisac, and A.D.Dragan]

Learning robot objective functions from human input has become increasingly important, but state-of-the-art techniques assume that the human's desired objective lies within the robot's hypothesis space. When this is not true, even methods that keep track of uncertainty over the objective fail because they reason about which hypothesis might be correct, and not whether any of the hypotheses are correct. We focus specifically on learning from physical human corrections during the robot's task execution, where not having a rich enough hypothesis space leads to the robot updating its objective in ways that the person did not actually intend. We observe that such corrections appear irrelevant to the robot, because they are not the best way of achieving any of the candidate objectives. Instead of naively trusting and learning from every human interaction, we propose robots learn conservatively by reasoning in real time about how relevant the human's correction is for the robot's hypothesis space. We test our inference method in an experiment with human interaction data, and demonstrate that this alleviates unintended learning in an in-person user study with a 7DoF robot manipulator.

Inference of constraints

Inferring and assisting with constraints in shared autonomy [CDC'16]
[Mehr et al.(2016)Mehr, Horowitz, and Dragan]

Our goal is to enable robots to better assist people with motor impairments in day-to-day tasks. Currently, such robots are teleoperated, which is tedious. It requires carefully maneuvering the robot by providing input through some interface. This is further complicated because most tasks are filled with constraints, e.g. on how much the end effector can tilt before the glass that the robot is carrying spills. Satisfying these constraints can be difficult or even impossible with the latency, bandwidth, and resolution of the input interface. We seek to make operating these robots more efficient and reduce cognitive load on the operator. Given that manipulation research is not advanced enough to make these robots autonomous in the near term, achieving this goal requires finding aspects of these tasks that are difficult for human operators to achieve, but easy to automate with current capabilities. We propose constraints are the key: maintaining task constraints is the most difficult part of the task for operators, yet it is easy to do autonomously. We introduce a method for inferring constraints from operator input, along with a confidence-based way of assisting the user in maintaining them, and evaluate in a user study.

Expression beyond goal configurations

Expressing objectives: Enabling Robots to Communicate their Objectives [RSS'17]
[Huang et al.(2017)Huang, Abbeel, and Dragan]

The overarching goal of this work is to efficiently enable end-users to correctly anticipate a robot's behavior in novel situations. Since a robot's behavior is often a direct result of its underlying objective function, our insight is that end-users need to have an accurate mental model of this objective function in order to understand and predict what the robot will do. While people naturally develop such a mental model over time through observing the robot act, this familiarization process may be lengthy. Our approach reduces this time by having the robot model how people infer objectives from observed behavior, and then it selects those behaviors that are maximally informative. The problem of computing a posterior over objectives from observed behavior is known as Inverse Reinforcement Learning (IRL), and has been applied to robots learning human objectives. We consider the problem where the roles of human and robot are swapped. Our main contribution is to recognize that unlike robots, humans will not be exact in their IRL inference. We thus introduce two factors to define candidate approximate-inference models for human learning in this setting, and analyze them in a user study in the autonomous driving domain. We show that certain approximate-inference models lead to the robot generating example behaviors that better enable users to anticipate what it will do in novel situations. Our results also suggest, however, that additional research is needed in modeling how humans extrapolate from examples of robot behavior.

Expressing constraints and capability: Expressing robot incapability [HRI'18, best paper nomination]
[Kwon et al.(2018)Kwon, Huang, and Dragan]

Our goal is to enable robots to express their incapability, and to do so in a way that communicates both what they are trying to accomplish and why they are unable to accomplish it. We frame this as a trajectory optimization problem: maximize the similarity between the motion expressing incapability and what would amount to successful task execution, while obeying the physical limits of the robot. We introduce and evaluate candidate similarity measures, and show that one in particular generalizes to a range of tasks, while producing expressive motions that are tailored to each task. Our user study supports that our approach automatically generates motions expressing incapability that communicate both what and why to end-users, and improve their overall perception of the robot and willingness to collaborate with it in the future.

Expressing future task plans: Generating plans that predict themselves [WAFR'16] [Fisac et al.(2016)Fisac, Liu, Harick, Hedrick, Sastry, Griffiths, and Dragan]

Collaboration requires coordination, and we coordinate by anticipating our teammates' future actions and adapting to their plan. In some cases, our teammates' actions early on can give us a clear idea of what the remainder of their plan is, i.e. what action sequence we should expect. In others, they might leave us less confident, or even lead us to the wrong conclusion. Our goal is for robot actions to fall in the first category: we want to enable robots to select their actions in such a way that human collaborators can easily use them to correctly anticipate what will follow. While previous work has focused on finding initial plans that convey a set goal, here we focus on finding two portions of a plan such that the initial portion conveys the final one. We introduce t-predictability: a measure that quantifies the accuracy and confidence with which human observers can predict the remaining robot plan from the overall task goal and the observed initial t actions in the plan. We contribute a method for generating t-predictable plans: we search for a full plan that accomplishes the task, but in which the first t actions make it as easy as possible to infer the remaining ones. The result is often different from the most efficient plan, in which the initial actions might leave a lot of ambiguity as to how the task will be completed. Through an online experiment and an in-person user study with physical robots, we find that our approach outperforms a traditional efficiency-based planner in objective and subjective collaboration metrics.

Expressing emotions: Teaching Robots to Span the Space of Functional Expressive Motion [IROS'22] [Sripathy et al.(2022)Sripathy, Bobu, Li, Sreenath, Brown, and Dragan]

Our goal is to enable robots to perform functional tasks in emotive ways, be it in response to their users' emotional states, or expressive of their confidence levels. Prior work has proposed learning independent cost functions from user feedback for each target emotion, so that the robot may optimize it alongside task and environment specific objectives for any situation it encounters. However, this approach is inefficient when modeling multiple emotions and unable to generalize to new ones. In this work, we leverage the fact that emotions are not independent of each other: they are related through a latent space of Valence-Arousal-Dominance (VAD). Our key idea is to learn a model for how trajectories map onto VAD with user labels. Considering the distance between a trajectory's mapping and a target VAD allows this single model to represent cost functions for all emotions. As a result 1) all user feedback can contribute to learning about every emotion; 2) the robot can generate trajectories for any emotion in the space instead of only a few predefined ones; and 3) the robot can respond emotively to user-generated natural language by mapping it to a target VAD. We introduce a method that interactively learns to map trajectories to this latent space and test it in simulation and in a user study. In experiments, we use a simple vacuum robot as well as the Cassie biped.

3 Impact

This project produced 17+ publications. 4 of them were best paper finalists, 3 of which at HRI, the premier human-robot interaction conference. It led to workshop talks at all the top robotics, HRI, and learning conferences (RSS, IROS, ICRA, HRI, NeurIPS, ICML, ICLR). During the course of the project and in part due to the work done as part of the project, the PI received a number of recognitions, including PECASE, TR35, Okawa, Sloan, IEEE RAS Early Career Award, IJCAI Early Career Spotlight. She was also invited for Keynotes at IROS, CoRL, and ICAPS.

References

[and. Krasheninnikov et al.(2019)and. Krasheninnikov, Alexander, Abbeel, and Dragan] R. Shah and. Krasheninnikov, J. Alexander, P. Abbeel, and A.D. Dragan. Preferences implicit in the state of the world. In *International Conference on Learning Representations (ICLR)*, 2019.

[Bajcsy et al.(2017)Bajcsy, Losey, O'Malley, and Dragan] A. Bajcsy, D. Losey, M. O'Malley, and A.D. Dragan. Learning robot objectives from physical human interaction. In *Conference on Robot Learning (CoRL)*, 2017. **(oral), acceptance rate 10%**.

[Bajcsy et al.(2018)Bajcsy, Losey, O'Malley, and Dragan] A. Bajcsy, D. Losey, M. O'Malley, and A.D. Dragan. Learning from physical human corrections, one feature at a time. In *International Conference on Human-Robot Interaction (HRI)*, 2018.

[Bhatia et al.(2020)Bhatia, Pananjady, Bartlett, Dragan, and Wainwright] K. Bhatia, A. Pananjady, P.L. Bartlett, A.D. Dragan, and M.J. Wainwright. Preference learning along multiple criteria: A game-theoretic perspective. In *Neural Information Processing Systems (NeurIPS)*, 2020.

[Bobu et al.(2018)Bobu, Bajcsy, Fisac, and A.D.Dragan] A. Bobu, A. Bajcsy, J. Fisac, and A.D.Dragan. Learning under misspecified objective spaces. In *Conference on Robot Learning (CoRL)*, 2018. **(invited to special issue)**.

[Bobu et al.(2020a)Bobu, Bajcsy, Fisac, and A.D.Dragan] A. Bobu, A. Bajcsy, J. Fisac, and A.D.Dragan. Quantifying hypothesis space misspecification in learning from human-robot demonstrations and physical corrections. In *IEEE Transactions on Robotics (TRO)*, 2020a.

[Bobu et al.(2020b)Bobu, Scobee, Satry, and Dragan] A. Bobu, D. Scobee, S. Satry, and A.D. Dragan. Less is more: Rethinking probabilistic models of human behavior. In *International Conference on Human-Robot Interaction (HRI)*, 2020b. **(best paper award)**.

[Bobu et al.(2021)Bobu, Wiggert, Tomlin, and Dragan] A. Bobu, M. Wiggert, C. Tomlin, and A.D. Dragan. Feature expansive reward learning: Rethinking human input. In *International Conference on Human-Robot Interaction (HRI)*, 2021. **(best paper finalist)**.

[Bobu et al.(2022)Bobu, Wiggert, Tomlin, and Dragan] A. Bobu, M. Wiggert, C. Tomlin, and A. D. Dragan. Inducing structure in reward learning by learning features. *International Journal of Robotics Research*, 2022.

[Chan et al.(2019)Chan, Hadfield-Menell, Srinivasa, and Dragan] L. Chan, D. Hadfield-Menell, S. Srinivasa, and A.D. Dragan. The assistive multi-armed bandit. In *International Conference on Human-Robot Interaction (HRI)*, 2019.

[Fisac et al.(2016)Fisac, Liu, Harick, Hedrick, Sastry, Griffiths, and Dragan] J. Fisac, C. Liu, J. Harick, K. Hedrick, S. Sastry, T. Griffiths, and A.D. Dragan. Generating plans that predict themselves. In *Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2016.

[Huang et al.(2017)Huang, Abbeel, and Dragan] S. Huang, P. Abbeel, and A.D. Dragan. Enabling robots to communicate their objectives. In *Robotics: Science and Systems (RSS)*, 2017. **(invited to special issue)**.

[Kwon et al.(2018)Kwon, Huang, and Dragan] M. Kwon, S. Huang, and A.D. Dragan. Expressing robot incapability. In *International Conference on Human-Robot Interaction (HRI)*, 2018. **(best paper award finalist)**.

[Lindner et al.(2021)Lindner, Shah, Abbeel, and Dragan] D. Lindner, R. Shah, P. Abbeel, and A.D. Dragan. Learning what to do by simulating the past. In *International Conference on Learning Representations (ICLR)*, 2021.

[Mehr et al.(2016)Mehr, Horowitz, and Dragan] N. Mehr, R. Horowitz, and A.D. Dragan. Inferring and assisting with constraints in shared autonomy. In *Conference on Decision and Control (CDC)*, 2016.

[Reddy et al.(2018)Reddy, Dragan, and Levine] S. Reddy, A.D. Dragan, and S. Levine. Where do you think you're going? inferring beliefs about dynamics from behavior. In *Neural Information Processing Systems (NeurIPS)*, 2018.

[Sripathy et al.(2022)Sripathy, Bobu, Li, Sreenath, Brown, and Dragan] A. Sripathy, A. Bobu, Z. Li, K. Sreenath, D.S. Brown, and A.D. Dragan. Teaching robots to span the space of functional expressive motion. In *International Conference on Intelligent Robots and Systems (IROS)*, 2022.