Towards an Information Theoretic Analysis of Human-Robot Interaction in Shared Autonomy

Thesis proposal

Deepak Edakkattil Gopinath
Department of Mechanical Engineering
Northwestern University
deepakgopinath@u.northwestern.edu

November 6, 2018

Abstract

Human-Robot Interaction (HRI) in the context of shared autonomy is a rich and complex phenomena. The effectiveness and usefulness of shared-control human-machine systems critically depends on the fluency and efficacy of human-robot interaction. Efficient HRI can lead to an improvement in joint task performance with higher user satisfaction and enhanced trust, all of which are desired characteristics of a joint human-machine system. From an engineer/systems designer's perspective, in order to achieve optimal performance the design of autonomy should adequately taken into account the richness, subtleties and complexity of the interaction between the human and the machine.

In this thesis proposal, I plan to propose a mathematical framework for human-robot interaction in the context of shared autonomy utilizing ideas from probabilistic graphical models and information theory. More specifically, the interaction between human and autonomy will be modeled as coupled perception-action loops unfolding in time using *causal Bayesian Networks*. Within this framework of causal Bayesian Networks, design of autonomy can be thought of as appropriately timed *interventions* at specific parts of model, with an intention to alter the bi-directional information flow between the human and machine. Using the proposed mathematical model, I will research three important problems that arise in HRI within shared autonomy, namely, a) learning b) inference and c) joint task performance. More specifically, I will focus on information theoretic analysis of how each of the above mentioned phenomena unfolds during task execution. The eventual goal is to utlize the proposed mathematical framework to inform the design of autonomy that will help *facilitate human learning*, *improve inference accuracy* and *enhance task performance*..

Contents

1 Introduction

Robots are ubiquitous in the modern-day society and have revolutionized the relationship between man and machine. Compared to a few decades ago, in the present day, robots have transitioned out of the rigid, structured and specialized industrial environments to the more rich, complex and unpredictable day-to-day human environments and have impacted diverse domains of human endeavor such as healthcare (medical, assistive and rehabilitation robotics), entertainment (musical robots) and home robotics **R**.

The impact is even more significant in the domain of assistive and rehabilitation robotics in which the potential to drastically enhance the quality of life for people suffering from motor impairments as a result of spinal cord or brain injuries is immense **R**. Devices such as smart wheelchairs, exoskeletons and assistive robotic arms can help to promote independence, boost self-esteem and help to extend mobility and manipulation capabilities of motor-impaired individuals and can revolutionize how they interact with society **R**.

The standard usage of these assistive machines, however, still relies on manual teleoperation by the human typically enacted through a control interface such as a joystick or a switch-based headarray **R**; that is, in such scenarios robots are not endowed with any intelligence and can be thought of as *passive* machines that function as extensions of human motor abilities. However, one of the most difficult conundrums is that greater the motor impairment of the user, the more limited the interfaces that are available for them to use. As a result, control of these machines can become extremely difficult due to the low dimensionality, sparsity and bandwidth of the control interfaces and are further exacerbated by the inherent complexity in robot dynamics and the physical limitations of the users **R**. In such cases, *robot autonomy*, the ability of robots to accomplish a task independently without requiring explicit instructions from a human, holds considerable promise as a tool to offset (and in some cases restore) the above-mentioned limitations. Advances in the fields of machine learning and artificial intelligence have helped to endow these assistive machines with better decision making and prediction capabilities while interacting with humans in real-world scenarios **R**. However, in literature there is a growing consensus that users of assistive technologies *do not* prefer to cede full control authority to the robotic partner during task execution **R**. Users, in general, like to have a more active role when interacting with an assistive robot **R**. In such cases, the introduction of *shared autonomy* seeks to find a middle ground between full teleoperation and autonomy by offloading only some aspects of task execution to the autonomy **R**.

In a shared autonomy system, the task responsibility is split between the user and autonomy with the aim of reducing human effort in accomplishing a task. Human-Robot Interaction (HRI) in the context of shared autonomy is a rich and complex phenomena **R**. The effectiveness and usefulness of shared-control human-machine systems critically depends on the quality and efficiency of human-robot interaction. That is, for robots and humans to work side-by-side and achieve joint goals and accomplish various tasks in a coordinated and cooperative manner, it is imperative that both parties understand each other, communicate and infer internal desires and intentions efficiently **R**. From an engineering perspective, design of appropriate kinds of autonomous behaviors for a shared-control system, therefore, needs to take into account the dynamics of human-robot interaction during the course of task execution **R**. This points to the need for rich mathematical frameworks that will model all the relevant variables and their interactions **R**.

Current research approaches for design of shared autonomy systems rely on various types of mathematical models and heuristics to solve different aspects of the problem independently and therefore suffer from generalizability across tasks, robotic platforms and various types of users. For my thesis proposal, I am motivated by the desire to develop a *unified* mathematical framework to analyze different aspects of human-robot interaction under a common umbrella in an attempt to shed light on the more *fundamental* and *low-level* descriptors.

To that end, I plan to propose a mathematical framework that models human-robot interaction in the context of shared autonomy, utilizing ideas from *probabilistic graphical models***R** and *information theory***R**. More specifically, the interaction will be modeled as coupled perception-action loops unfolding in time using *causal Bayesian Networks* **R**. The nodes in the network will represent the different variables (both latent and observed) that are relevant for the model and the edges represent the probabilistic influence they have on each other**R**. In an attempt to quantify the fluency, transparency and cooperation of human-robot interaction heavy emphasis will be placed on analyzing the *information flow* between the nodes in the network. Within this proposed framework of causal Bayesian networks, design of autonomy can be thought of as appropriately timed *interventions***R** that has the potential to alter bidirectional information flow between human and autonomy. Our hypothesis is that *information flow* is a more fundamental and low-level descriptor of joint system performance that system designers should focus on when designing autonomous behaviors. Using the proposed model, I propose to address three main subproblems relevant to shared autonomy namely, *learning*, *inference* and *task performance*.

The first research question (**RQ1**) that I will address in my work is *how can autonomy help humans learn robot dynamics better*. When a human and machine interact in a shared autonomy setting, both parties are continually learning about each other**R**. For example, for novice users, with practice their familiarity with the device increases and they learn about the dynamics of the control interface and the robot**R**. The initial forward (and inverse) dynamics model that the user maintains internally during task execution might be drastically different from the true dynamics**R**. Due to learning effects, the internal model will tend towards the true model. However, the learning strategies that humans adopt need not always be optimal, for example, users might not sample the state and action space in an efficient and exhaustive manner and therefore can erroneously extrapolate dynamics between different regions of the workspace**R**. Therefore, autonomy can play the role of a *teacher* and help the human in skill acquisition and provide appropriate guidance during the learning process**R**.

Inherent limitations of the control interface and motor impairments, however, can possibly put an upper bound to skill level that can be acquired. In such scenarios, the need for autonomy becomes inevitable. However, any successful assistive robotic system needs to have a good idea of the user's needs and intentions. That is, *user intent inference* is a necessary and crucial component to ensure proper assistance**R**. Therefore, the second research question (**RQ2**) that I will address in my thesis is *how can autonomy be designed so that inference becomes more accurate*. Typically, the user's internal state (desires, goals and intentions) is latent (if not fully, partially) from autonomy's perspective**R**. It has to be noted that inference is not a unidirectional phenomena. For example, from the users' perspective the internal logic with which autonomy helps them is not always explicitly known and therefore needs to be inferred as well. User satisfaction and acceptance heavily depends on the user's understanding of how the autonomy works**R**. In this thesis, I plan to utilize the proposed mathematical model to reason about and shape the information flow from the user's internal state to autonomy to improve the inference accuracy.

In addition to facilitating learning (**RQ1**), and improving inference accuracy (**RQ2**), autonomy has to work in conjunction with the human to perform the task optimally. Therefore, the third and final research question (**RQ3**) that I hope to tackle in this thesis is *how to design autonomy to ensure optimal task performance* **R**. Typically, both subjective (user satisfaction, acceptance, trust) and objective metrics (task completion time, number of mode switches) equally inform the optimality criteria **R**. Rather than focusing on the above-mentioned metrics independently, in this work we will focus on optimal bidirectional information flow between the human and autonomy. Our hypothesis is that optimization of information flow between the autonomy and human will result in better communication of latent states. This will likely lead to enhanced cooperation and mutual understanding as a result of which the desired outcomes (better task performance, improved user satisfaction) will naturally emerge.

In summary, in this proposed thesis I intend to develop a mathematical framework to model human-robot interaction in shared autonomy and plan to research solutions to the questions presented above (RQ1, RQ2 and RQ3). I am motivated by the need to develop a unified theoretical framework for shared autonomy. This work will be the first to treat information content and flow as the key components to understand the dynamics of interaction between human and autonomy in a shared autonomy setting. More importantly, this work proposes a fundamentally different way of thinking about autonomy; one in which *autonomy is a exogenous intervention that alters the information flow in a coupled perception-action loop to bring about desired outcomes*.

2 Human-Robot Interaction in Shared Autonomy

2.1 Mathematical Models for Shared Autonomy

2.2 Aspects of HRI

2.3 Causal Bayesian Networks for Modeling HRI

2.4 Information Theory and Flow in CBNs

3 Proposed Work

3.1 Inference

Different types of inference. Make inference easier by becoming more legible, transparent. Bidirectional intent inference. We are focused on human to machine. DEvelop information thoeretci framework for goal disambiguation.

3.1.1 Related Work

3.1.2 Experiment Design

3.2 Learning

Guided Active Learning. Where the autonomy is the guide. Human do active learning (tinkering) on their own. Autonomy, plays the role of a teacher/guide and guides the active exploration so that the rate of learning is higher for the human.

3.2.1 Related Work

3.2.2 Experiment Design

3.3 Task Performance

3.3.1 Related Work

3.3.2 Experiment Design

4 Timeline

Timeline	Work	Progress
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	completed
Nov. xxxx	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	ongoing
Jan. xxxx	Thesis writting	
Feb. xxxx	Thesis defense	

Table 1: Plan for completion of my research

Thus, I plan to defend my thesis in XXX XXXX.

References