Research Summary and Plans

Mohan Sridharan

I am broadly interested in cognition and control in robots[1] and humans. My primary research interests include knowledge representation and reasoning, cognitive systems, machine learning, and control systems. I develop algorithms and architectures that: (a) *represent*, *reason*, and *act* reliably and efficiently with *qualitative* and *quantitative* descriptions of commonsense domain knowledge and uncertainty; (b) *learn interactively* and *cumulatively* from noisy multimodal sensor cues; and (c) enable designers to *understand* the behavior of robots and humans, and to establish that this behavior *satisfies desired properties*. I pursue an integrated cognitive systems approach, jointly addressing the underlying challenges by drawing on and deepening our understanding of human cognition and control. Furthermore, I develop such algorithms to enhance autonomy in non-robotics application domains.

Knowledge representation and reasoning: Robots often have to make decisions over long time horizons based on prior knowledge and the information extracted from noisy multimodal data streams. This includes qualitative descriptions of commonsense knowledge, e.g., default statements such as "books are usually in the library" that hold true in all but a few exceptional circumstances, and quantitative descriptions that associate probabilities with the information extracted from sensor observations, e.g., "I am 90% certain the robotics book is in the library". I have developed architectures that leverage the complementary strengths of **non-monotonic logics** and **probabilistic reasoning** to represent and reason with these descriptions. These architectures encode **cognitive theories**, e.g., of *intention* and *affordance*, and are based on transition diagrams of the domain at two resolutions, with a fine-resolution transition diagram defined formally as a **refinement** of a coarse-resolution transition diagram. For any given goal, non-monotonic logical reasoning with commonsense knowledge at the coarse-resolution provides a sequence of abstract actions. Each abstract action is implemented and executed as a sequence of concrete actions by automatically **zooming** to and reasoning with the relevant part of the fine-resolution diagram, using probabilistic models of uncertainty (if available) and suitable methods, e.g., hierarchical partially observable Markov decision processes. The fine-resolution outcomes are automatically propagated to the coarse-resolution for subsequent reasoning. I am currently extending this methodology to explore new representations and processing mechanisms, using the learning algorithms described below to identify relevant representations for different tasks.

Interactive Learning and Explainable Agency: My research poses the acquisition of previously unknown knowledge as an **interactive learning** problem, with the information needed for knowledge revision obtained from humans or the environment. Although humans can provide rich information about tasks and the domain, humans are unlikely to have the time and expertise to interpret raw sensor inputs or to provide comprehensive feedback. My research enables robots to **learn associations** between **multimodal cues** (e.g., images, verbal descriptions), building rich object representations to ground words describing object attributes and relations between objects. The learned representations also help identify inconsistencies (e.g., unexpected outcomes, ambiguous task description), and to solicit human feedback by automatically constructing and **posing clarification questions** based on need and availability.

Learning from observing or interacting with the domain is challenging because it is difficult to obtain a large number of labeled training examples in complex domains. In my architectures, the robot reasons with domain knowledge to automatically guide learning to objects and events relevant to tasks at hand. Specifically, **non-monotonic logical reasoning** with domain knowledge triggers learning when necessary, e.g., when unexpected outcomes are observed. Reasoning also guides the use of different **machine learning** algorithms, e.g., *relational reinforcement learning*, *inductive learning*, or *deep learning*, to automatically

[1] I use the terms "robot" and "agent" interchangeably in my research.

learn and revise relevant domain knowledge, e.g., previously unknown actions and axioms governing domain dynamics, from a small number of training examples. Much of this work is illustrated and evaluated in the context of vision-based planning, scene understanding, and question answering tasks.

The interplay between representation, reasoning, and learning is leveraged to implement a **theory of explanation** and achieve key functional capabilities of **explainable agency**. In particular, the robot is able to provide on-demand relational descriptions as *explanation* or justification of decisions made before, during, or after planning and execution by considering alternative choices. It is able to run *mental simulations* to identify and present information at a suitable level of abstraction in response to different types of questions (e.g., causal, contrastive, counterfactual), and to use adaptive **theory of mind** models to communicate information such that it makes contact with human concepts such as beliefs and goals.

Robot Manipulation and Control: Similar to my work on reasoning and learning, the algorithms and architectures I develop for dexterous robot manipulation are inspired by findings in human motor control. My recent work has explored **changing-contact manipulation** tasks, which require the robot to make and break contact with different objects and surfaces. These tasks are representative of many robot (and human) manipulation tasks, and are characterized by piecewise continuous interaction dynamics that can damage the robot and the domain objects. In a departure from existing work, my recent architecture enabled a robot manipulator to incrementally learn **forward (predictive) models** of the end-effector measurements. The prediction errors are used to automatically revise the forward models, guide the use of transition-phase controllers, and to and vary the impedance (i.e., stiffness, damping) parameters of **hybrid force-motion controllers** in order to follow a desired motion pattern. More recent work has developed an architecture that combines parameterized human muscle-tendon models with variable impedance controllers toward achieving **adaptive control of upper-limb prostheses**.

Human-agent/Multiagent Collaboration: The challenges described above are more pronounced when we consider robots collaborating with other robots or humans. I am particularly interested in **ad hoc teamwork**, i.e., enabling an agent to collaborate with others without prior coordination, which is representative of many practical applications. State of the art methods include a data-driven component to model the behavior of other agents and optimize the behavior of the ad hoc agent based on a long history of prior experiences. My research, on the other hand, combines knowledge-based reasoning and data-driven learning, embedding the principle of **ecological rationality** and an *algorithmic model of heuristics* in a refinement-based architecture that focuses on **adaptive satisficing** instead of optimization. As a result, each *ad hoc agent* automatically identifies and uses relevant heuristic methods to rapidly build predictive models of the behavior of others (teammates, opponents) from limited examples. The agent then reasons with prior knowledge and these learned behavior models to make decisions and adapt to unforeseen changes.

Practical Applications: The architectures and algorithms I develop are grounded in practical applications in the context of **robots assisting humans** in offices and homes, or in applications such as agricultural automation and automated exploration. In addition, I have designed and adapted algorithms and architectures to address automation challenges in non-robotics domains such as climate informatics, transportation, agricultural irrigation management, and software testing For example, in collaboration with *Auckland Transport*, I adapted machine learning algorithms for **short-term traffic prediction** on the motorways in Auckland (NZ). Also, in collaboration with researchers at the *U.S. Department of Agriculture*'s Agricultural Research Service, I developed methods to accurately estimate reference evapotranspiration values for **irrigation scheduling**, and to estimate crop yield from satellite images. Another research project developed computational models for **predicting extreme weather events** and for understanding the relationships between global and regional climate models under climate change scenarios.

To summarize, my research jointly investigates fundamental knowledge representation, reasoning, control, and learning challenges in the context of robots and humans. I pursue an integrated cognitive systems approach to address these challenges, motivated by the goal of enabling widespread use of robots and software systems that assist and collaborate with humans. I seek to pursue my research objectives in a vibrant multi-disciplinary atmosphere in collaboration with colleagues and students. Additional details and publications are available online: `https://www.cs.bham.ac.uk/~sridharm/`