Understanding How Humans Teach Robots

Tasneem Kaochar

Department of Computer Science, University of Arizona
kaochar@email.arizona.edu

Abstract. Robots and other intelligent devices capable of carrying out highly complex procedures are becoming ubiquitous in the home and workplace. However, changing the behavior and capabilities of these devices typically requires direct programming by specially trained engineers. While machine learning (ML) algorithms offer the allure of allowing machines to improve their knowledge and behavior from experience, ML algorithms still require considerable expertise to use in practice. To bridge this gap, human-instructable computing seeks to develop intelligent devices that can be taught by natural human instruction. Our research focus is on developing methods for non-expert humans to teach complex behaviors to autonomous agents by accommodating natural forms of human teaching. Currently, most systems for human-robot teaching allow only one mode of teacher-student interaction (e.g., teaching by demonstration or feedback), and teaching episodes have to be carefully set up by an expert. To understand how we might integrate multiple, interleaved forms of human instruction into a robot learner, we performed a behavioral study in which untrained humans were allowed to freely mix interaction modes to teach a simulated robot (secretly controlled by a human) a complex task. We found that teaching styles varied considerably but can be roughly categorized based on the types of interaction, patterns of testing, and general organization of the lessons given by the teacher. Analysis of transcripts showed that human teachers often give instructions that are nontrivial to interpret and not easily translated into a form useable by ML algorithms. In particular, humans often use implicit instructions, fail to clearly indicate the boundaries of procedures, and tightly interleave testing, feedback, and new instruction. Our study contributes to a better understanding of human teaching patterns, highlights the challenges of building an initial automatic teacher interpretation system using ML algorithms and makes specific recommendations for future human-robot interaction systems.