Active Sensing in Complex Multiagent Environments (Doctoral Consortium)

Adam Eck
Department of Computer Science and Engineering
University of Nebraska-Lincoln
aeck@cse.unl.edu

ABSTRACT
In this research, we focus on active sensing solutions to address challenging properties in complex environments, such as uncertainty, partial observability, non-stationarity, and limited resources. We describe our ongoing contributions, focusing on sensing for both individual agents and cooperating teams. We also outline how we are applying our research to two real-world applications: personal assistants and intelligent survey systems.

Categories and Subject Descriptors
I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence – *intelligent agents, multiagent systems*

Keywords
Active Sensing; Observer Effect; PBRS; Information Sharing

1. INTRODUCTION
Multiagent systems (MAS) are commonly applied to many real-world applications, such as robotics, wireless sensors networks, cyber-physical systems, and human-computer interactions. One commonality to these applications is the *complexity* of the environment, including difficult properties such as uncertainty, partial observability, noise, non-stationarity, real-time constraints, limited resources, and multiple actors. Each property constrains an agent's ability to gather information, make informed decisions, and prudently act to accomplish its goals. Therefore, an agent must address these properties either implicitly or explicitly to operate properly and enable the entire system to achieve the desired emergent behavior and achieve system-wide goals.

In this research, we address the challenging properties of complex environments in the agent's **sensing** activities, since without good information, a rational agent cannot make good decisions or act appropriately. We develop solutions using active sensing, whereby an agent makes intelligent meta-level decisions about the information gathered for its reasoning. We focus on both *single agent sensing*, independent of other agents, as well as information sharing and sensing adaptation between *many coordinating agents*. We apply these solutions to several real-world problems, including personal assistants and intelligent survey systems.

2. ACTIVE SENSING
To properly control sensing in order to address the challenging properties inherent in complex environments, we rely on **active sensing** (also known as **active perception**). From this perspective, an agent *actively* reasons about its sensing behavior, selecting sensing actions to maximize observational value (e.g., accuracy, uncertainty reduction) and minimize negative impacts (e.g., costs, resulting environment changes, limited resource use), instead of *passively* receiving whatever observations happen to be produced by the environment [13]. Using active sensing explicitly balances the benefits and costs of sensing actions, causing an agent to *proactively* aim to maximize its sensing performance, and not only *reactively* rely on suboptimal observations.

Within the context of active sensing, we have studied a novel complex environment property called the **Observer Effect** [6], caused by using ***stateful* resources** during agent sensing. Specifically, some resources used by agents have an internal state that determines their behavior upon use (e.g., wireless network bandwidth, human user frustration and cognitive load). The mere act of sensing with these resources to gather information about the environment (e.g., monitoring the wireless network through special transmitted packets, interrupting a user to discover her preferences) can ultimately change the state of the resource, and thus directly impact the quality and quantity of observations. Therefore, special care must be taken to not corrupt the state of the resource and distort the information gathered during sensing.

To manage stateful resources and mitigate the Observer Effect, we model active sensing using a popular (e.g., [1, 12]) approach: the partially observable Markov decision process (POMDP [9]). Using this model, an agent can explicitly reason about the state of both the resources used (which impact sensing outcomes) and its knowledge (which is improved through sensing), as well as the changes to these states based on performing sensing actions. For the reward maximized by the agent, we model the improvement in the value of the knowledge held by the agent for its particular tasks, which reflects both (1) the improvement in knowledge from high quality observations, and (2) the distortion in knowledge caused by bad resource states. Since this improvement might not be known a priori, we use reinforcement learning (RL) to acquire a model of knowledge refinement to guide sensing.

Through this work, we recognized that rewarding based on changes in agent knowledge (or beliefs) is not traditional for active sensing POMDPs, which tend to focus instead on costs of sensing actions or rewards for individual (hidden) states. After searching the literature, we discovered similar belief-based rewards [1], such as measuring the uncertainty in an agent's belief state resulting from an action. To better understand and exploit different types of rewards, we have been studying (1) the *advantages and disadvantages* of each type of reward for active sensing [4]: traditional rewards [9] (beneficial for considering sensing costs), belief-based rewards [1] (beneficial for considering belief improvement), and hybrid combinations [1], and (2) methods to combine different types of rewards to achieve *theoretical performance guarantees.*

To achieve such theoretical guarantees, we were inspired by a related solution in RL: **potential-based reward shaping** (PBRS) [2, 10]. PBRS shapes an agent's traditional reward function to include additional information describing the value to an agent of transitioning from one state to another (in terms of likely earning large future rewards). Thus, PBRS combines two (or more) types of information in a shaped reward function and has the additional

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013),* Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

advantage of still optimizing the agent's traditional rewards in spite of this addition. We have produced the first extension of PBRS to online POMDP planning [7] in order to both (1) combine multiple perspectives in the agent's rewards (e.g., action costs and belief-state improvement), as well as (2) improve short horizon planning by *implicitly* guiding the agent towards large future rewards beyond its planning horizon. It can be shown that this extension retains the theoretical guarantees from its use in RL (e.g., [2, 10]), and thus is an appropriate way to combine multiple types of rewards. Furthermore, this work is beneficial for POMDPs in general (even beyond active sensing) as it can better maximize the agent's cumulative rewards by including information (e.g., belief certainty) not traditionally considered in POMDPs. We are currently studying how to better select which types of rewards to combine using PBRS, based on complex environment properties.

We have also recently begun investigating the application of active sensing to *cooperating agents* in MAS. Specifically, we have focused on the **large team information sharing** problem (LTIS, e.g., [8, 11]). In this problem, a large team of agents (e.g., 1000 agents) work together to form consistent, accurate beliefs about some environment phenomenon, but only a small subset (e.g., 5% of the agents) can directly observe the environment, whilst all other agents must rely on shared information to form beliefs. Our main contribution to this area thus far is studying **non-stationary** phenomenon that change values over time [5] (e.g., time-varying user preferences in large systems of human-agent interactions). We established that non-stationarity greatly complicates the problem due to an institutional memory emergent behavior, where the team cannot overcome initial beliefs in spite of the weight placed on new information during Bayesian belief updating (an important focus with stationary phenomenon, e.g. [8, 11]). We are developing new solutions to this problem, such as (1) change detection and response within local neighborhoods, and (2) forgetting outdated information using belief decay. We are also interested in further contributing to this area by considering additional complex environment properties, such as the costs of performing sensor observations and communication between agents, real-time constraints on belief convergence throughout the team, and the Observer Effect from multiple agents directly monitoring the phenomenon using shared stateful resources.

3. APPLICATIONS

We are currently working to apply our research on active sensing (both individual agent and multiagent) to two real-world applications: (1) personal assistant agents and (2) intelligent survey systems. First, within our Adaptive Knowledge Assistants framework [3], personal agents assist human knowledge workers to produce new knowledge from information in domains such as scientific research, citizen science, and education. These agents use active sensing to both (1) discover their users' preferences and goals while mitigating the Observer Effect caused by distracting the user during her tasks, and (2) autonomously manage the users' informational needs, such as finding novel sources of information, matchmaking users to share information, and automating routine tasks used to gather information for the users. We are applying this framework to an intelligent wiki system (the Written Agora) used for computer supported, collaborative learning in a number of settings: from topic-based articles and discussions in computer science courses to an online repository for a student-produced journal documenting group-based biology experiments.

Second, we are also working to apply active sensing to intelligent survey systems. Here, we are interested in developing agents responsible for controlling and adapting surveys online as they are presented to human users in order to improve the end-user's experience and the quality and quantity of information collected from users. This work is being applied to both web-based personal surveys, as well as computer-assisted time diary interviews.

Beyond these two applications in human-computer (and human-agent) interactions, this research has potential broader impacts in domains such as robotics (e.g., search and rescue robots exploring unknown spaces), wireless sensor networks (e.g., limited network bandwidth monitoring and control), and cyber-physical systems (e.g., controlling limited numbers of sensors to form consistent beliefs across many devices)

4. ACKNOWLEDGMENTS

This research has been supported by a NSF Graduate Research Fellowship (DGE-054850), a U.S. Department of Education GAANN Fellowship (P200A040150), as well as grants from the NSF (DBI-0743783, SBES-1228937), and has been partially completed utilizing the Holland Computing Center of the University of Nebraska.

5. REFERENCES

[1] Araya-Lopez, M., Buffet, O., Thomas, V., and Charpillet, F. 2010. A POMDP extension with belief-dependent rewards. *Proc. of NIPS'10*.

[2] Devlin, S. and Kudenko, D. 2012. Dynamic potential-based reward shaping. *Proc. of AAMAS'12*.

[3] Eck, A. and Soh, L.-K. 2012. Adaptive Knowledge Assistants. *Proc. of HAIDM'12*, 107-114.

[4] Eck, A. and Soh, L.-K. 2012. Evaluating POMDP Rewards for Active Perception (Extended Abstract). *Proc. of AAMAS'12*.

[5] Eck, A. and Soh, L.-K. 2013. Dynamic Facts within Large Team Information Sharing (Extended Abstract). *Proc. of AAMAS'13*.

[6] Eck, A. and Soh, L.-K. 2013. Observer Effect from Stateful Resources in Agent Sensing. *Autonomous Agents and Multi-agent Systems*, 26(2), 202-244.

[7] Eck, A., Soh, L.-K., Devlin, S., and Kudenko, D. 2013. Potential-Based Reward Shaping for POMDPs (Extended Abstract). *Proc. of AAMAS'13.*

[8] Glinton, R., Scerri, P., and Sycara, K. 2010. Exploiting scale invariant dynamics for efficient information propagation in large teams. *Proc. of AAMAS'10*, 21-28.

[9] Kaelbling, L.P., Littman, M.L., and Cassandra, A.R. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101, 99-134.

[10] Ng, A.Y., Harada, D., and Russell, S. 1999. Policy invariance under reward transformations: Theory and application to reward shaping. *Proc. of ICML'99*, 278-287.

[11] Pryymak, O., Rogers, A., and Jennings, N.R. 2012. Efficient opinion sharing in large decentralized teams. *Proc. of AAMAS'12*.

[12] Spaan, M.T.J., Veiga, T.S., and Lima, P.U. 2010. Active cooperative perception in networked robotic systems using POMDPs. *Proc. of IROS'10*, 4800-4805.

[13] Weyns, D., Steegmans, E., and Holvoet, T. 2004. Towards active perception in situated multi-agent systems. *Applied Artificial Intelligence*, 18, 867-883.