Leveraging Online Virtual Agents to Crowdsource Human-Robot Interaction

Nick DePalma
MIT Media Lab
20 Ames Street
Cambridge, MA 02139
ndepalma@media.mit.edu

Sonia Chernova
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
soniac@cs.wpi.edu

Cynthia Breazeal
MIT Media Lab
20 Ames Street
Cambridge, MA 02139
cynthiab@media.mit.edu

ABSTRACT

Robots require a broad range of interaction skills in order to work effectively alongside humans. They must have the ability to detect and recognize the actions and intentions of a person, produce functionally valid and situationally appropriate actions, and engage in social interactions through physical cues and dialog. However, social interactions with one of today's robots will quickly become one-sided and repetitive, even after just a few minutes due to its shallow depth of knowledge and experience. This problem exposes weaknesses in the underlying traditional approaches that aim to pre-code responses to a limited number of inputs. We propose the use of crowdsourcing as a tool for the development of social robots that allow for rich, diverse and natural human-robot interaction. To enable crowdsourcing at a massive scale, we describe a newly implemented system that uses online virtual agents to collect data and then leverages the resulting corpus to train our robot behavior system for use on a real world task.

Author Keywords
Crowdsourcing, Human-Robot Interaction

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: Human Factors

General Terms
Design, Robotics

INTRODUCTION

Socially situated agents typically demand an incredible amount of information for a sufficiently impressive and useful interaction from the humans perspective. Robotic social agent research has revealed a number of important topics that require data driven approaches. Implementation of these social interaction techniques rely more on design than science: transparency, contingency, saliency, group dynamics, motivation, and turntaking have not fully been understood but have many adaptations; while other interaction subsystems can leverage domains with support from larger communities in AI and perception: language processing, perception, navigation, and group dynamics. Understanding these new complex, contextual interactions may could utilize large amounts of data. Unfortunately all of these domains are highly interconnected and may require them all to be trained simultaneously to correlate information between these domains.

Learned constructivist approaches toward artificial intelligence is nothing new. Drescher [5] explored this work in detail regarding virtual agents, and related work in robotics [3] has mirrored this approach with physical agents and makes similar arguments regarding long term constructivist approaches toward the AI demands of socially intelligent agents. To build large corpora of data, the robotics community has been struck by a predicament; how do we teach a robot from "infancy" to "adulthood"? We are interested in pushing the community to deliver more complex, interrelated data to train robots in the real world, in effect accelerating the development of a robot's cognitive architecture.

We introduce and discuss the merits of our newly implemented system that uses online games to crowdsource social interactions for physical systems. Our preliminary results are promising and we're hoping that the Human Robot Interaction community will benefit from utilizing our data, which we intend to release to the public, to train new innovative models of behavior.

LOOK TO CROWDSOURCING

Crowdsourcing offers an answer to our problems. Teaching and training a robot from scratch demands a corpus larger and more complex than can be gathered in small incremental interactions. Moreover, crowdsourcing can be far more variable in content and context. Our hypothesis is that by leveraging the wisdom of the crowds, we can begin to train our robots to produce deeper interactions with humans. Osentoski et. al. [10] has already begun using closed loop control and maze-solving with crowds in the loop and has shown positive results.

Human Robot Interaction with humanoid robots is a brand of robotics that demands that the robots satisfy all of the ex-

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

pectations that science fiction has set forth. Some of these demands include a proficiency with dialog or language in general, a theory of mind, an ability to perform tasks; an ability to be autonomous in new situations; an ability to learn from these new experiences; and some would say an ability to understand, recognize and show emotion which is sometimes framed as a prerequisite to interesting dialog and interaction. While this is not the place to debate the intricacies of such interaction decisions, this can be seen as a subset of just some of the more in demand and critical pieces of sufficient interaction design.

Crowdsourcing may offer a great way of gathering much needed datasets that are needed to train our models for interesting behavior. Our previous work [2] offers a glimpse of what we intend to do with crowdsourcing but we believe we can do more and leverage the crowds to produce even more knowledgeable robots. We believe that human-robot interaction can benefit from leveraging crowds to obtain training data for following topics: kinematics, labeling, language, skill acquisition, and behavior design.

Kinematics

Kinematics is an area that has already had success. Sorokin et. al. [11] have used crowds to source information on how to grasp objects by utilizing camera or depth information and the aid of human collaborators to triangulate a mesh and determine the best way to grasp such an object. Complex tool use also demands that object kinematics be known. Learned tool use has been explored in previous work by Katz et. al. [4]. The dynamics of new tools are complex and may demand the help of humans to determine points of articulation that could be useful for robots in the real world.

Labeling

Symbol-grounding is an important step in language acquisition. Simply knowing the name of objects without building the behavior required to interact with a human is a giant step. Simply labeling the objects that are reconstructed using standard and depth (RGBD) cameras would aid in simple object based interaction between a robot and a human. Sorokin [7] has already leveraged standard cameras to dynamically augment the known database of objects using human online collaborators in real time.

Language acquisition

Apart from classification and grounding, language acquisition in general can benefit from a more complex corpus. Dialog and group discussions within contextual tasks are usually limited to a relatively small number of topics. We argue that these small number of topics can be covered within a large number of small interactions between players in a virtual space. These observed discussions can be used to train speech recognition systems to cover potential contextual discussions, as well as capturing non verbal cues to mimic that may be absent from prototypical language learning data corpora.

Skill acquisition

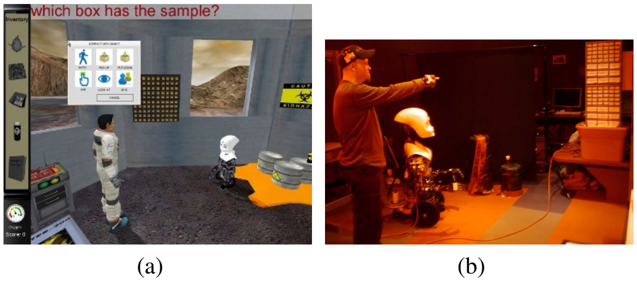
Task learning is sometimes viewed as the primary role of robots. While there are many ways to design a robot, most robots are built around performing a single task and others are built generically to perform or be trained to perform a small number of tasks. These more general robots can be more flexible if they have the ability to acquire new skills on demand through robot learning and planning. Goal oriented skill-acquisition taxonomies can be seen as a naturally hierarchical and structured. This has led to scaffolded approaches to robot skill acquisition but are typically preprogrammed with the basic skills required to learn a small number of tasks. Experiments on skill acquisition over long-term interactions and the basic skills demanded of more complex scaffolded tasks needs to be explored. We believe that crowdsourcing offers a unique opportunity to learn, structure and scaffold more complex goal oriented skills that cannot be reasonably taught to a robot directly. Crowdsourcing may provide a more canonical approach to bootstrapping the robot for basic skills before passing the robot into new environments. This topic is not new but is important to note in the context of the position.

Behavior Design

Non-verbal communication is a major concern for human robot interaction. Gesticulation, attention mechanisms, pointing, disposition and other such psychologically inspired behaviors become increasingly important for natural interaction [6, 1]. Models of these behaviors have been described in philosophical literature but these models lack algorithmic grounding. For example, saliency and focus are known factors in attention but what to pay attention to and in what context are not as easily known. Crowd sourcing offers a mechanism, we believe, to give us insight into what constitutes natural behaviors. By posing different contextual scenarios, virtual agents can illuminate how to naturally use these mechanisms.

CROWDSOURCING SOCIAL BEHAVIORS USING ONLINE GAMES

In our current work, we are exploring the use of online games as a means of generating large-scale data corpora for human-robot interaction research. We present a system in which action and dialog models for a collaborative task involving a person and a robot are learned based on a reproduction of the task in an online multiplayer game. Similar to projects


which box has the sample?

(a) (b)

Figure 1. (a) A screenshot of the *Mars Escape* game showing the action menu and dialog between the players. (b) A picture of the *Mars Escape* environment recreated in the real world.

such as Games with a Purpose [13] and the ESP Game [12], our goal is to make work fun in order to harness the computational power of Internet users. Our work is inspired by the Restaurant Game project [8, 9], in which data collected from thousands of players in an online game is used to acquire contextualized models of language and behavior for automated agents engaged in collaborative activities.

The goal of our research is to enable a robot to perform a collaborative task with a human by leveraging a corpus of example interactions collected in an online game. For testing, we have selected a general search and retrieval task in which the robot and human must work together to collect multiple objects. The task has no strictly assigned social roles, however, the domain is developed to encourage collaborative behaviors such as action synchronization, sequencing and dialog.

Data Collection

We collect data using a custom-made online two-player game called *Mars Escape*. The game records the actions and dialog of two players as they take on the roles of a robot and an astronaut performing a collaborative task on Mars. During the game, the players are able to navigate in the environment, manipulate objects using a number of predetermined actions (e.g., look at and pick up) and communicate with each other through in-game text-based chat (see Figure 1(a)). At the completion of the game, the players are given individual scores based on the number of items collected and the time required to complete the mission. Players are also asked to complete a survey evaluating their gaming experience and the performance of their partner. Players are asked to rate how much they enjoyed working with the other person, as well as to speculate on whether the other character was controlled by a person or an AI.

Data Analysis

During the first three months of the release of the game we captured data from over 500 two-player games. We utilize this data to develop action and dialog models for the robot character. In our analysis, we are interested in identifying sequences of behaviors that represent typical player behavior.

To model what a "typical" interaction might look like, we utilize the Plan Network representation developed by Orkin and Roy for the Restaurant Game [8]. A Plan Network is a statistical model that encodes context-sensitive expected patterns of behavior and language. Given a large corpus of data, a Plan Network provides a mechanism for analyzing action ordering and for visualizing the graphical structure of action sequences.

Figure 2(a) shows the full graph of all action transitions recorded for the robot player. The graph is so complex, that it is impossible to extract much meaningful information from its representation. However, crowdsourcing enables us to look at this data in the context of a large population and to eliminate behavior sequences that were taken only by a few individuals (have a low likelihood) and can therefore be assumed to be anomalous or unusual (for example, attempting to climb furniture). Figure 2(b) presents the graph in which all transitions that were observed in less than 2% of logs are eliminated. The graph is significantly simpler than the full Plan Network, highlighting the power of crowdsourcing this kind of data. While many players choose to deviate from the norm at some point in their gaming experience, aberrant interactions wash away statistically when compared to the larger number of examples of typical behavior.

Finally, Figure 3 shows a *Wordle*[1] representing the diversity of dialog phrases used in player interaction. The size of the font reflects the frequency with which that phrase occurs. The diversity and size of this space far surpasses that of most pre-coded dialog systems for human-robot interaction. This database can be used by the robot both as a reference to look up the input received from the human, as well as to generate spoken output.

Evaluation using a Physical Robot

In the final phase of this project, we are evaluating the action and dialog models learned from the crowdsourced data corpus in a real-world variant of the Mars Escape task. Evaluation is being performed at the Boston Museum of Science, where museum visitors are recruited to perform the task in collaboration with our autonomous MDS robot *Nexi* (Figure 1(b)). The MDS robot platform combines a mobile base with a socially expressive face and two dexterous hands that provide the capability to grasp and lift objects. The robot is equipped with a biologically-inspired vision system that supports animate vision for shared attention to visually communicate the robot's intentions to human observers.

CONCLUSION

By leveraging large corpora of various crowdsourced data, we believe that Human Robot Interaction will become more data driven, will become more rigorous in its approach to scaffolding interesting behaviors and may aid in comparing interactions by utilizing shared corpora of data. By opening up interactions sourced from crowds from various simulations, various models can be trained to bootstrap the primary robotic architecture that will be used as a baseline for further learning by demonstration techniques. We are encouraged by the success of others and the potential success of our own experiment and are optimistic about the future of crowdsourced data to train our models and to help direct the field to build metrics that may be useful in benchmarking our findings against one another.

Acknowledgments

We'd like to acknowledge the help of Jeff Orkin whose insights, inspiration, and advice has been critical to the direction of this project. We'd also like to thank Elisabeth Morant whose vital effort to clean and train the models have been indispensable. This project was funded by the Office of Naval Research.

REFERENCES

[1] http://www.wordle.net/

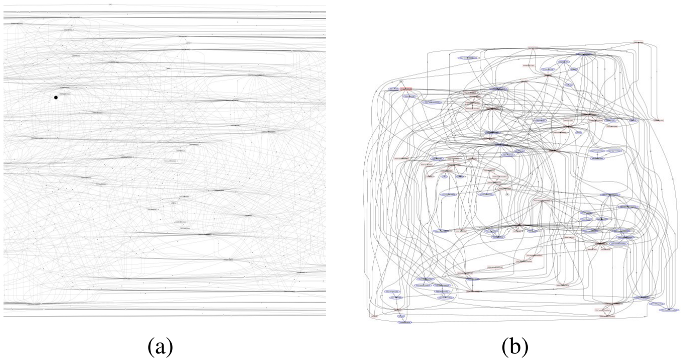

(a) (b)

Figure 2. a) Part of a Plan Network visualizing typical behavior for the astronaut role across 350 games. b) The same Plan Network in which all transitions occurring in fewer than 2% of the logs have been omitted.

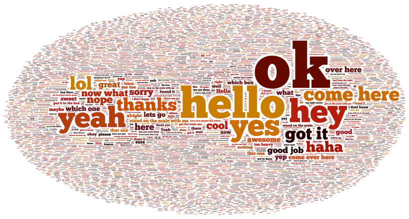


Figure 3. A *Wordle* representing the most commonly observed phrases, as well as the diversity of inputs.

1. Breazeal, C., Kidd, C., Thomaz, A., Hoffman, G., and Berlin, M. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, IEEE (2005), 708–713.
2. Chernova, S., Orkin, J., and Breazeal, C. Crowdsourcing HRI Through Online Multiplayer Games. In *Papers from the AAAI Fall Symposium, Dialog with Robots*, AAAI (2010).
3. Dautenhahn, K., and Billard, A. Bringing up robots or the psychology of socially intelligent robots: From theory to implementation. In *Proceedings of the third annual conference on Autonomous Agents*, ACM (1999), 366–367.
4. Dov Katz, Yuri Pyuro, O. B. Learning to manipulate articulated objects in unstructured environments using a grounded relational representation. In *Proceedings of Robotics: Science and Systems IV* (Zurich, Switzerland, June 2008).
5. Drescher, G. *Made-up minds: a constructivist approach to artificial intelligence*. The MIT Press, 1991.
6. Fong, T., Thorpe, C., and Baur, C. Collaboration, dialogue, human-robot interaction. *Robotics Research* (2003), 255–266.
7. Giles, J. Robot uses human ingenuity in bid for self-reliance. *The New Scientist 203*, 2726 (2009), 22.
8. Orkin, J., and Roy, D. The restaurant game: Learning social behavior and language from thousands of players online. *Journal of Game Development* (2007).
9. Orkin, J., and Roy, D. Automatic learning and generation of social behavior from collective human gameplay. In *AAMAS* (2009), 385–392.
10. Osentoski, S., Crick, C., Jay, G., and Jenkins, O. Crowdsourcing for closed-loop control. *Neural Information Processing Systems, NIPS 2010 Workshop on Computational Social Science and the Wisdom of Crowds* (2010).
11. Sorokin, A., Berenson, D., Srinivasa, S., and Hebert, M. People helping robots helping people: Crowdsourcing for grasping novel objects. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, IEEE (2010), 2117–2122.
12. von Ahn, L., and Dabbish, L. Labeling images with a computer game. In *CHI*, ACM (New York, NY, USA, 2004), 319–326.
13. von Ahn, L., and Dabbish, L. Designing games with a purpose. *Communications of the ACM 51*, 8 (2008), 58–67.

Bio of the Author

Nick DePalma is a student in the MIT Media Lab. He holds a BS and an MS in computer science from the Georgia Institute of Technology with focuses in graphics, user interface, artificial intelligence, and robotics. He has worked in the game and computer vision industry and intends to build helpful, entertaining, and enjoyable cooperative robots for those that need it most.