Avoiding Robot Faux Pas: Using Social Context to Teach Robots Behavioral Propriety

Cory J. Hayes, Maria F. O’Connor, and Laurel D. Riek
Dept. of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
{chayes3,moconn13,lriek}@nd.edu

ABSTRACT
Contextual cues strongly influence the behavior of people in social environments, and people are very adept at interpreting and responding to these cues. While robots are becoming increasingly present in these spaces, they do not yet share humans’ essential sense of contextually-bounded social propriety. However, it is essential for robots to be able to modify their behavior depending on context so that they operate in an appropriate manner across a variety of situations. In our work, we are building models of context for social robots, that operate on real-world, naturalistic, noisy data, across multi-context and multi-person settings. In this paper, we discuss one aspect of this work, which concerns teaching a robot an appropriateness function for interrupting a person in a public space. We trained a support-vector machine (SVM) to learn an association between contextual cues and the reaction of people being interrupted by a robot across three different contexts. Overall, our results are promising, and further work on integrating context models into social robots could lead to interesting and impactful findings across the HRI community.

Categories and Subject Descriptors
H.1 [Models and Principles]: User/Machine Systems

Keywords
social robotics; human-robot interaction; propriety

1. INTRODUCTION
Context influences people’s behavior in human social environments such as workplaces, social events, and home settings [3, 5, 6]. These types of environmental influences lead groups to develop a shared internal sense of propriety which links behaviors with situational parameters. Bar’s neurological model of visual context, represented in Fig. 2, proposes an explanation of how this link naturally occurs [1]. In this model, in-depth and computationally intensive processing, such as person and object recognition, occurs simultaneously with top-down low-resolution processing where contextual associations are made based on the scene as a whole.

While much research in robotics and computer vision focuses on in-depth processing in object recognition scenarios, little work has been done in taking the quicker, top-down approach. In our work, we use thin-slicing methods to learn one particular contextual association, an appropriateness function, based on real-world data collected from a social, mobile robot. This binary function defines a mapping from a sensed context to an appropriateness value for some robot behavior.

Social propriety for robotics is becoming an important issue in the field of HRI [2]. In order for robotic systems to be affordable, adaptable, and acceptable to users, they must be able to conform to their environment without requiring developers to anticipate and develop for all possible contexts. Our work attempts to integrate context in the process of robots learning behavioral propriety.

Our work addresses two research questions. The first is “Can high level contexts be automatically differentiated?” since robots must be able to discriminate between contexts before adapting their behavior. If this is possible, we then ask “Can we use contextual cues to learn an binary appropriateness function for some behavior?” In this work, we propose an appropriateness function to teach a physically embodied robot behavioral propriety after interacting with people in naturalistic contexts.
2. METHODS

We conducted an study in which we used a modified Turtlebot and custom Android application to interrupt and initiate interactions with participants (see Fig. 1). The robot’s height was raised to human-level so that it could more easily interact with participants audio-visually via the tablet interface. The robot recorded visual data using a Microsoft Kinect, and audio data using a digital voice recorder. To prevent the robot from toppling over, its motion was fully controlled by a Wizard via line-of-sight. The Wizard also remotely launched the Android application (which was fully autonomous; it just did not easily integrate with ROS). Because our robot is usually capable of full autonomy for these tasks, we did not simulate operator error [4].

After spotting a potential participant, the Wizard navigated the robot toward the person. Once the robot was in position, the Wizard triggered the interaction screen on the tablet, which consisted of a text-captioned audio prompt of “Hello. Is this a good time to bother you?” If the participant responded “Yes”, the application recorded the response and ended the interaction after thanking the participant. In the case of a “No”, the robot spoke briefly about the university before ending the interaction in the same manner.

In order to gather data to train an appropriateness function, our robot explored three types of social contexts on our college campus: study areas, dining areas, and lobby areas, across both the student center and library.

Our experiment consisted of two phases. The first phase (learning) involved gathering data from interruptions to train our appropriateness function. The second phase (validation) tested the accuracy of our appropriateness function when the robot was once again placed in these locations.

Each interruption was represented by 10 seconds of data from both the recorded video and audio prior to the interruption. We used one key frame per second to represent the visual component of our audio-visual feature vector, and used each second of audio to represent the other. We then used principal component analysis (PCA) to reduce the dimensionality of each vector. (PCA reduces dimensionality in a lossy way and is analogous to the fast, low-dimension processing system in the brain, predicted by Bar’s model [1].) We performed classification using 10-fold cross-validation and a linear SVM.

3. RESULTS

We recorded 60 interactions for the learning phase. We first determined if the features we extracted were sufficient for classifying scenes by context. We achieved an overall classification accuracy of 87.17% for context. Using the same feature vectors as before without knowledge of context, we then trained and tested our appropriateness function for interruption; this resulted in a classification accuracy of 80.5%.

For the second phase, we performed two tests to validate our approach to learning appropriate behavior. The first test used our original data for training and 109 interactions of new data for testing, using the same parameters as before. This resulted in a near-chance classification accuracy (52.11%), which we suspected was in part due to the kernel and parameters used for selection, as well as potential over-fitting and the unseen data problem. Consequently, we then used a polynomial kernel to repeat learning, achieving an accuracy of 85.33%. This was similar to the initial phase, though the validation accuracy dropped to 58.07%.

4. DISCUSSION

While results from the learning stage appeared promising, results from the validation stage suggest that context is too large of a problem space to learn from quickly. These results shed some computational light on the need for a dual-processing system in the brain, predicted by Bar’s model [1]. Though the improvement after using a polynomial SVM kernel was not enough to make a large difference, we can see that it is possible to develop more accurate models if the system is able to change parameters, though this would require the automation of the machine learning process and further analysis to determine the optimal classifier.

Context and behavior are indeed linked, and our attempt to use contextual features to classify audiovisual scenes based on propriety was able to predict both the categorical context and appropriateness function for the behavior of interruption. Our initial results show that broad contexts can be learned from extremely noisy data. One clear motivation that our work provides is the consideration of combining both low-resolution and detailed processing streams. However, we first need to learn the associations between these broad contexts and particularly robust identifiable objects and actions.

It is also worth noting all of our data was naturalistic, meaning our results are closer to ecological validity than those based on experiments performed in laboratory settings. This is important for the development of social robotic systems intended to act independently in the “real-world” [5].

5. REFERENCES


Figure 2: Example of Bar’s visual context model.