Non verbal behavior modeling for Socially Assistive Robots

- Socially assistive robotics (SAR)

> Socially assistive robotics (SAR) is a subfield of robotics that aims to design, construct, and evaluate robots that help people through social interactions (Feil-Seifer and Matarić

> Efficient, intuitive human-robot communication is critical to SAR. People perform much of their communication nonverbally, using behaviors like eye gaze and gesture to convey mental state, to reinforce verbal communication, or to augment what is being said (Argyle 1972). Though these nonverbal behaviors are generally natural and effortless for people, they must be explicitly designed for robots. As SAR
>
> cation, robot behavior must follow human expectations. If robots generate social behavior that is outside of the established communication norms, people will be confused or reject the robot interaction outright. Therefore, any approach to designing social behaviors for robots must be informed by actual human behavior.

- Focus of research

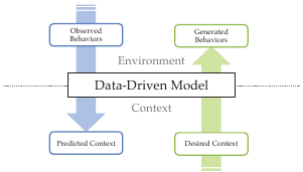
> Our research focuses on improving human-robot interaction in SAR applications by modeling the complex dynamics of human communication. We do so by building models of human-human nonverbal communication and implementing these models to generate socially appropriate and communicative behavior for socially assistive robots. These data-driven models allow us to design robots that match people's existing nonverbal communication use.

- Demand for data-driven robot behavior models

> Some researchers have begun to address this need for data-driven robot behavior models. For example, researchers have modeled conversational gaze aversions (Andrist et al. 2014) and gesture during narration (Huang and Mutlu 2013) based on analysis of human-human pairs. While these studies show promising advances in data-driven models of robot behavior, none of them deals directly with socially assistive applications, which require monitoring of—and feedback from—the interaction partner.

- Model

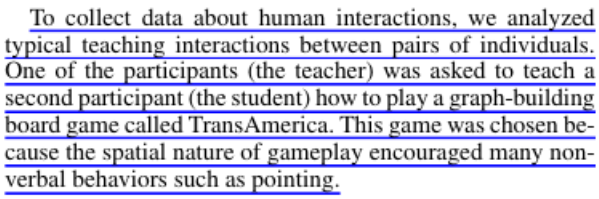
> Unlike previous work, our model is *bidirectional*, enabling both the *prediction* of a user's intent given observed nonverbal behaviors, and the *generation* of appropriate nonverbal



> There are three steps in the process of designing a data-driven generative behavior model: 1) collect data on nonverbal behaviors during human-human interactions, 2) train a predictive computational model with the human-human interaction data, and 3) develop a generative model for robot behaviors driven by the computational model from step 2.

This is similar to our multi-modal system in that we also collect data (hand gestures) from users and train a network to recognize the gestures.

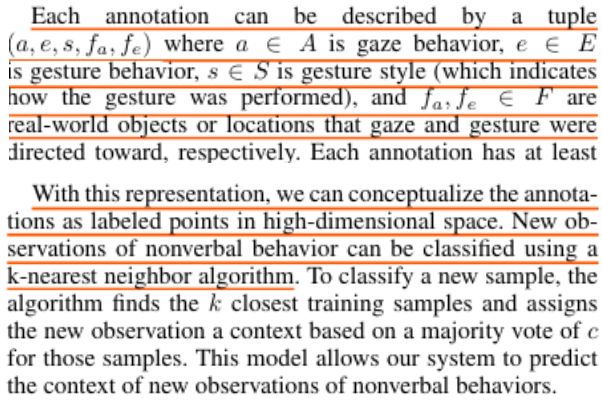
- Data collection



To collect data about human interactions, we analyzed typical teaching interactions between pairs of individuals. One of the participants (the teacher) was asked to teach a second participant (the student) how to play a graph-building board game called TransAmerica. This game was chosen because the spatial nature of gameplay encouraged many nonverbal behaviors such as pointing.

This data was then manually annotated (for non verbal behavioral features)

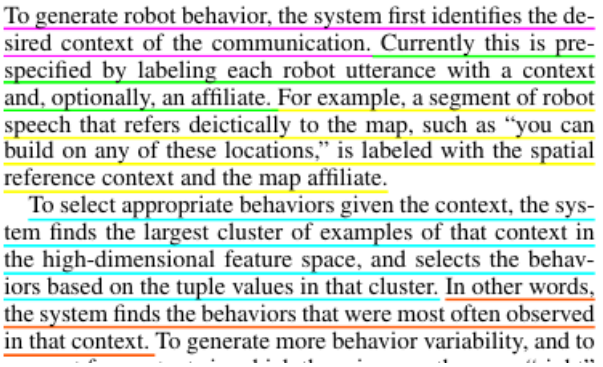
- Classification



Each annotation can be described by a tuple (a, e, s, f_a, f_e) where $a \in A$ is gaze behavior, $e \in E$ is gesture behavior, $s \in S$ is gesture style (which indicates how the gesture was performed), and $f_a, f_e \in F$ are real-world objects or locations that gaze and gesture were directed toward, respectively. Each annotation has at least

With this representation, we can conceptualize the annotations as labeled points in high-dimensional space. New observations of nonverbal behavior can be classified using a k-nearest neighbor algorithm. To classify a new sample, the algorithm finds the k closest training samples and assigns the new observation a context based on a majority vote of c for those samples. This model allows our system to predict the context of new observations of nonverbal behaviors.

- Robot behavior generation



To generate robot behavior, the system first identifies the desired context of the communication. Currently this is prespecified by labeling each robot utterance with a context and, optionally, an affiliate. For example, a segment of robot speech that refers deictically to the map, such as "you can build on any of these locations," is labeled with the spatial reference context and the map affiliate.

To select appropriate behaviors given the context, the system finds the largest cluster of examples of that context in the high-dimensional feature space, and selects the behaviors based on the tuple values in that cluster. In other words, the system finds the behaviors that were most often observed in that context. To generate more behavior variability, and to

This paper is similar to ours in the sense that we also try to predict context (but also commands) based on non-verbal input (gestures). Where it differs is of course the type of none-verbal input (they use gaze, gestures, context and objects) as well as they attempt as to generate robot behavior given a context (bi-directional)