

# Personalizing crowdsourced human-robot interaction through curiosity-driven learning

Phoebe Liu  
Figure Eight Technology Inc.  
U.S.A

Malcolm Doering  
Advanced Telecommunications Research  
Institute International  
Japan

Dylan F. Glas  
Futurewei Technologies  
U.S.A

Takayuki Kanda  
Department of Social Informatics  
Kyoto University  
Japan

Dana Kulić  
Department of Electric and Computer  
Engineering, University of Waterloo  
Canada

Hiroshi Ishiguro  
Department of Systems Innovation  
Osaka University  
Japan

**Abstract**—Learning human-robot interaction logic from crowdsourced data is a promising approach for generating robot behaviors, but behaviors learned only from offline data can sometimes become predictable and “robotic”. For example, a shopkeeper robot might always perform the same action when repeatedly encountering the same interaction state, resulting in monotonous, boring interactions for a customer. In order to maintain engagement as a robot interacts repeatedly with a user over the long term, it is important to adapt its behavior to that user. We developed a robot whose behavior is driven by curiosity, which first learns high-level dialog and spatial behavior patterns from offline examples of human-human interaction. Then, during live interactions, it chooses among appropriate actions according to its curiosity about the customer’s expected behavior, continually updating its predictive model to learn and adapt to each individual. We present a case study where the curious robot adapts to an engaged and unengaged customer, tailoring its actions in real time in order to explore and satisfy its curiosity about the customers’ individual differences.

**Keywords**— *Data-driven social interaction; curiosity-based learning*

## I. INTRODUCTION<sup>1</sup>

Thanks to the availability and the ease of crowdsourced data in recent years, it is possible to train conversational robots to reproduce such high-level behavior patterns without modeling the underlying cognitive decision processes, as demonstrated in prior work for robots in the role of a shopkeeper [2], an assistant [3], a bartender [4], and a storyteller [10].

While existing data-driven approaches enable a robot to generate socially appropriate behaviors, such learning-based techniques only learn from a fixed collection of behaviors [2]–[7] and can become repetitive over time, causing user engagement to decrease after novelty wears off.

For example, imagine a scenario where a shopkeeper engages a customer in a camera shop. The shopkeeper may ask what sort of camera the customer is interested in but get no response. At this point, the shopkeeper, having already tried the most promising strategy to engage the customer, will be driven

by his *intrinsic motivation* to try different actions, which could possibly lead to a favorable outcome.

Inspired by our own sense of curiosity, we propose a data-driven approach to enable a robot to continuously adapt to a user’s behavior during online interaction while remaining task-relevant and socially appropriate (e.g. not saying “goodbye” when the customer enters the shop). Using our approach, the robot first learns task-appropriate behaviors *a priori* [7], [13] from crowdsourced interaction data, and then tailors its actions in real time in order to satisfy its curiosity about the customers’ individual differences. Rather than fully replicating the human curiosity mechanism, our aim is to emulate curiosity in the robot for the purpose of creating a more humanlike interaction.

## II. RELATED WORK

A common first step to data-driven HRI is the acquisition of task-specific interaction data. Many researchers have used crowdsourcing platforms, such as Figure Eight<sup>2</sup>, to quickly acquire data of many types (including text, audio data, video annotations, multi-lingual data, etc.) at relatively low cost to rapidly bootstrap and prototype dialog agents [5], [8]–[10]. Likewise, training data can also be acquired through collaborative games [3], [11], remote web users, and real human interaction data [6], [12], [13]. Our work complements these approaches, considering crowd-based data collected directly from human-human interaction in a physical environment; however, we focus on the robot continuing to learn based on intrinsic motivation, to better adapt to each user over time.

Some studies have demonstrated user-adaptive techniques to improve interactions with different users [14]. Gordon et al. present a robotic tutoring system that aims to enhance the child’s learning experience by maintaining knowledge on the child’s reading level and periodically evaluates and updates it using an Active Learning technique [15]. Similarly, Parundekar and Oguchi [16] demonstrated how a driver’s preferences can be modeled through implicit and explicit feedback to better personalize the result for Points of Interest. Our work also aims to continuously adapt to an individual in real-time, but for a social robot interacting through speech and locomotion.

<sup>1</sup> This paper summarizes the techniques presented in [1], currently under submission. For further details and a user study evaluation, please refer to [1].

<sup>2</sup> <https://www.figure-eight.com/>

### III. DATA COLLECTION

#### A. Scenario

To observe typical interaction patterns, we set up a camera shop environment in an 8m x 11m experiment space with three camera models at different locations. For each interaction, one shopkeeper participant interacted with one customer participant. An interaction example for our scenario is shown in Fig. 1.

#### B. Sensors

We recorded the participants’ speech and position data as they interacted with each other. A sensor network [17] consisting of 16 ceiling-mounted Microsoft Kinect RGBD sensors, estimates the position and body orientation of each person based on point cloud data. We captured the participants’ speech using handheld smartphones through a push-to-talk app based on the Google speech recognition API.

#### C. Participants

We recruited fluent English speakers as participants for the role of the customer. They had varied levels of knowledge about cameras. We employed a total of 18 participants (13 male, 5 female, average age 32.8, s.d. 12.4). Since we hoped to obtain a diverse set of behaviors, we chose two different participants to role-play as the shopkeeper. They had very different interaction styles, one with a more outgoing personality (male, age 54) and another with a quieter disposition (female, age 25).

#### D. Procedure

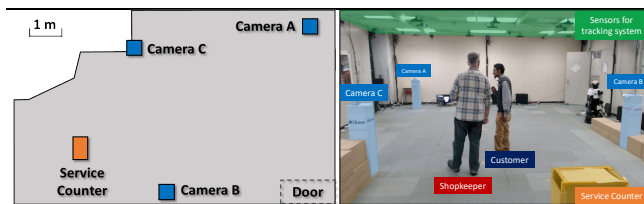
Participants were encouraged to act naturally and focus discussion on the features listed on the camera spec. sheets (8 to 10 features per camera). Customer participants were encouraged to play with the cameras, browse the shop, or ask camera-related questions. For variety, they played advanced or novice camera users in different interactions. Shopkeeper participants were instructed to begin interactions at the service counter, be polite, and behave according to their role (e.g. greetings and farewells, letting the customer browse, answering questions, or introducing products when appropriate). As the example in Fig. 2 shows, participants used a variety of fillers (e.g. “you know”, “like”) and backchannels (e.g. “I see”) in their utterances. Therefore, we believe our setup elicited reasonably natural behaviors.

Each shopkeeper interacted with 9 different customer participants. Each customer role-played 24 interactions (12 as advanced and 12 as novice) for a total of 216 interactions. 27 interactions were removed (16 due to technical failures and 11 due to customers that did not follow instructions). None of the interactions are repeated. In total, we collected 405 interactions, with 4061 shopkeeper utterances and 4115 customer utterances.

### IV. PROPOSED TECHNIQUE

#### A. Overview

In order to develop a curious robot that continues to learn during interaction, the robot first needs to learn social rules observable in the human-human data. For example, sales staff in a shop might develop routine techniques for presenting products to customers, and they might respond to typical questions with similar answers each time a new customer asked them. So, we first trained an *Appropriateness Learner*, using feature vectors extracted from the training data, which constrains the robot to a



(Customer enters the shop and shopkeeper approaches)  
 C: I'm looking for a camera that is stylish, affordable, and easy to use. Do you have one of those?  
 S: Depends on what sort of pictures you take, we have two very stylish affordable cameras, one is \$68 and one is over \$500.  
 C: Oh I see, I think I might be looking to get the one that costs 68 dollars, can you please show me?  
 S: (Present Nikon) Sure, just over here this camera takes beautiful pictures but it's designed to be point-and-shoot, catch the moment.  
 C: Can you explain to me about the preset modes?  
 S: It has 18 different modes where you can tell it what sort of conditions you are in ...  
 (Customer asks a few more questions about camera features and shopkeeper answers)

Figure 1. An example interaction observed in our data collection.

subset of possible behaviors that it can explore in a particular situation.

Next, we developed a *Curiosity Learner* to assign a curiosity score to each possible robot action. We model curiosity as the drive to minimize the variance of the prediction error of the consequence of the robot’s actions. Similar to the shopkeeper, a customer will generally behave within the norm of certain interaction patterns, which can be used as a prior for the *Curiosity Learner*. However, due to individual differences, different customers may react very differently to a given robot action, and it is these differences to which a curious robot must tailor its interaction dynamics. Thus, during live interaction, we are only interested in updating the *Curiosity Learner*.

As previously mentioned, the robot’s actions should be governed by both social appropriateness and curiosity, hence we applied a *behavior utility* function that combines these two factors. The robot executes the action with the highest *utility*, which balances social appropriateness and curiosity. Below, we describe briefly the details of the individual components of the curiosity-based learning systems, which is illustrated in Fig. 2.

**Abstraction and representation:** We abstracted continuous streams of captured sensor data into common behavior patterns (i.e. spatial formations and spoken utterances) in order to learn robot behaviors effectively despite the large variation of natural human behavior and noisy input from the sensor system. Using data processing and abstraction techniques from previous studies [2], the input vector consists of a sentence embedded using Latent Semantic Analysis and the participant’s abstracted motion state. The training target for our *learners* is a discrete value, obtained by hierarchical clustering of all the feature vectors into  $K$  clusters. Each discrete value is a concatenation of a typical utterance ID (e.g. ID 5) and a target spatial formation (e.g. *present Nikon*).

**Appropriateness Learner:** First, to learn the social appropriateness of a robot action, we trained a feed-forward multilayer perceptron (MLP) neural network, which has the ability to map the relative importance of the input vectors to a discrete training target. Our training data for the neural network is composed of  $(h(t), \hat{r}(t))$  action pairs, where  $h(t) \in \mathbb{R}^m$  is the

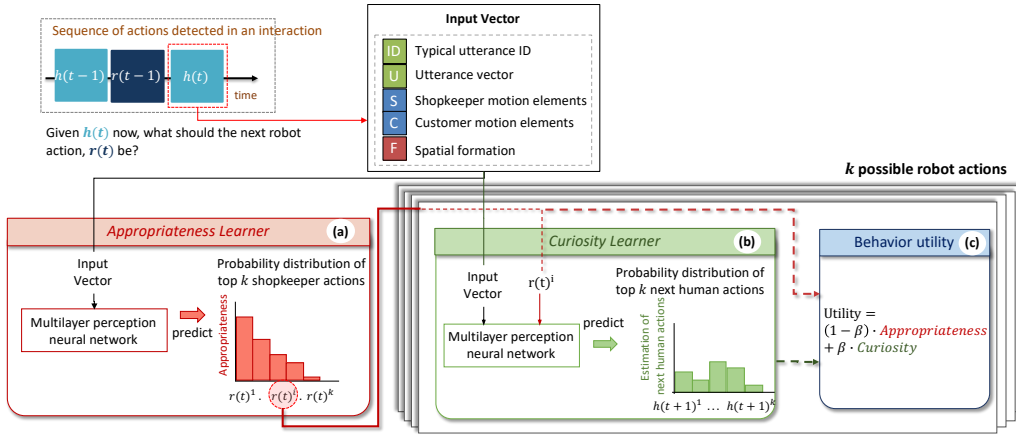


Figure 2. Details of our system, both the *Appropriateness Learner* and *Curiosity Learner* are triggered when a human action (e.g. utterance or silence) is detected. (a) The *Appropriateness Learner* learns a set of  $k$  socially appropriate robot actions, and for each robot action; (b) we query the *Curiosity Learner* to output a measure for *curiosity*; (c) finally a *utility* measure is calculated.

human action input vector and  $\hat{r}(t) \in \{0,1\}^K$  is a target robot shopkeeper action, where  $K$  is equal to the total number of robot actions obtained from clustering. That is, if  $\hat{r}(t)^i = 1$ , observation  $h(t)$  maps to robot action  $i$ .

Based on the results from previous studies, we can interpret the neural network as learning a measure of how *appropriate* each robot action is [18]:

$$\text{Appropriateness} = p(r(t)^1), \dots, p(r(t)^K) \quad (1)$$

During online interaction, we want to constrain the robot to the top  $k$  most appropriate behaviors predicted by the neural network for a particular situation, among which the robot can freely explore using the *Curiosity Learner*.

**Curiosity Learner:** We model the curiosity of a robot action as trying to minimize the variance of the prediction error [19], i.e. the robot is curious about those actions for which it is uncertain how the customer will respond, and less curious about actions for which it is confident it can predict what the customer will do next. Similar to the *Appropriateness Learner*, we first learn an initial estimation of commonly observable customer actions, by applying another MLP neural network.

Considering a sequence of alternating actions  $(h(t), r(t), h(t+1))$ , the training input for the neural network is  $(h(t), r(t))$ , and the training target is the discretized value,  $\hat{h}(t+1)$ . The neural network learns a probability distribution over the set of human actions in the next timestep,  $p(h(t+1)^1), \dots, p(h(t+1)^K)$ , where  $K$  is the total number of discrete human actions. Finally, only the top  $k$  most likely subsequent customer actions were used. For our case,  $K=800$  and  $k=5$ .

To measure the uncertainty of the customer's next action, we calculate the entropy of the probability distribution that is output by the neural network [20]. Previous computational models have also incorporated such uncertainty-based strategies, generating biases toward actions or states that have high entropy [21]. A high entropy value means that the robot is unsure what the customer will do as a result of its own action, while a low entropy value means that the robot is fairly confident of what the customer will do next. The robot is then encouraged to take actions that result in states that are deemed surprising, i.e. where the robot is unsure what the customer will do next.

Thus, the *curiosity* measure is the normalized entropy of the probability distribution:

$$\text{Curiosity} = \frac{-\sum_{i=1}^K p(h(t+1)^i) \log(p(h(t+1)^i))}{\ln(K)} \quad (2)$$

**Behavior utility:** For each potential robot action, a behavior utility function is evaluated, which combines the factors of *social appropriateness* and *curiosity*:

$$\text{Utility} = (1 - \beta) \cdot \text{Appropriateness} + \beta \cdot \text{Curiosity} \quad (3)$$

where  $\beta$  is a tuning parameter that is adjustable. A high  $\beta$  biases the robot to be more curious while a low  $\beta$  biases the robot to be more socially appropriate during interaction.

**Action selector:** To select a behavior for a robot, the behavior utility function is evaluated for each of the potential actions the robot can perform. The action selector then executes the discrete robot action with the highest utility, consisting of a typical utterance and a target spatial formation.

### B. Adaptation to individuals

As the robot continues to interact with a customer, it should come to better understand how the customer will respond to its actions, and thus tend to be less curious about actions it has taken previously. To reflect this new observation in the *Curiosity Learner*, during live interaction, we update the weights of the neural network in the *Curiosity Learner*. Backpropagation is used to modify the synaptic weights of the internal (hidden) and output layers of the neural network [38], by trying to minimize the loss between the target and the predicted value. In this way, the input-output mapping of the neural network can be dynamically updated to reflect new observations.

To update the weight of the neural network, the newly observed human action is first mapped to an action cluster,  $\hat{h}(t)$ , using the nearest neighbor algorithm. Then, this action is used as a target for backpropagation with the cross-entropy loss function. To control how quickly the neural network learns the observed human action, we backpropagate the newly observed interaction data through the neural network over several epochs until the cross-entropy loss is below a certain threshold,  $th$ . This allows the recently observed human action to immediately become the most likely prediction for that prompt.

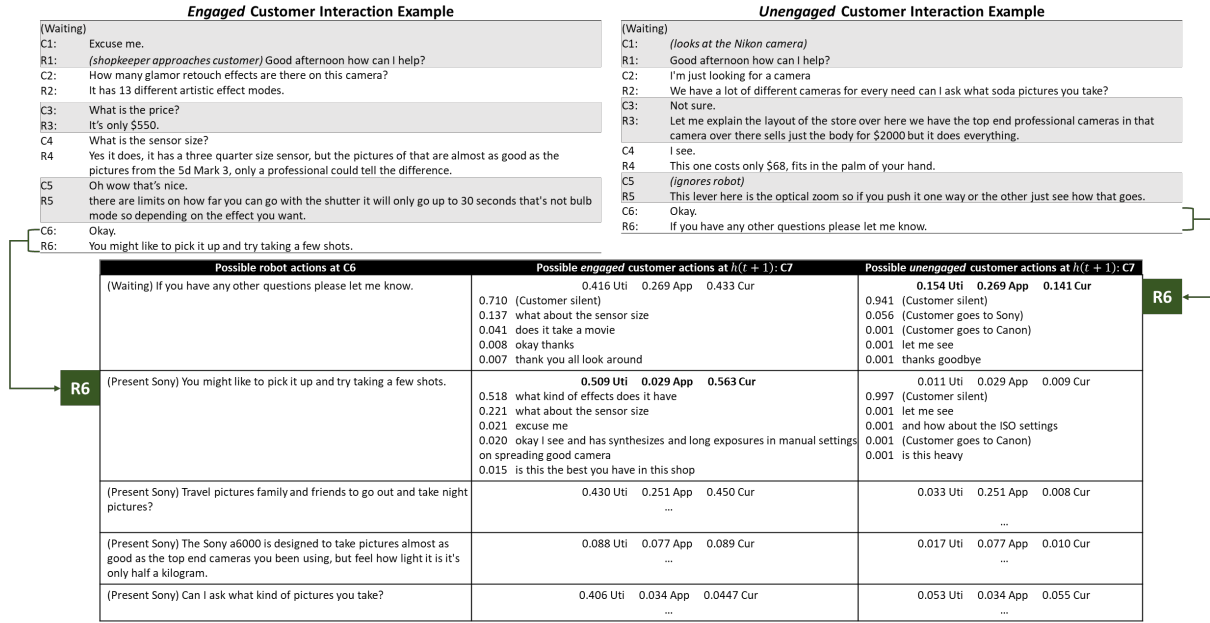


Figure 3. An example of the curiosity-driven robot. Uti, App, and Cur are the respective utility, appropriateness and curiosity values. The probability distribution of the next customer action (e.g. Prob C2),  $h(t+1)$  is also shown. For brevity, the predicted customer actions are only shown for the relevant robot actions.

### C. Model Parameters

The architectures of both neural networks are the same, consisting of an input layer, followed by three leaky rectified hidden layers, and a softmax output layer. The input to the *Appropriateness Learner* is the human action vector of dimension and the input to *Curiosity Learner* consists of both a human and a robot action vector with total dimension. Each hidden layer consists of 800 neurons.

Both neural networks were trained using momentum-based mini-batch stochastic gradient descent, with a batch size of 128, a learning rate of 0.0005, and a momentum coefficient of 0.9. Normalized initialization [22] was used to initialize the neural network. The network was trained to minimize the cross-entropy loss for 2000 epochs between the observed target action and the predicted action for the entire training set.

### V. CASE STUDY

Here we present an interaction example of how the curious robot adapts to engaged and unengaged customers in Fig. 3 using the proposed approach. Both customers perform the same action, but because the *Curiosity Learner* has adapted based on each customer’s previous behaviors, it is able to provide differing actions, tailored to their individual differences.

Fig. 3 (left) shows an example interaction for an engaged customer, who is asking the robot many questions. In action C6, the customer responds to the robot with a backchannel, “Okay.” Based on the customer’s previous actions, the *Curiosity Learner* has learned that he is likely to continue the conversation with any of a number of actions. So, the robot responds proactively, “You might like to pick it up and try taking a few shots.”

In contrast, the right side of Fig. 3 shows an example interaction of an unengaged customer. The robot tries offering information or asking questions, but is answered with short, disinterested responses (e.g. “Not sure” or ignored by the customer). Finally, when the customer responds with “Okay”,

the *Curiosity Learner*, having learned that the customer is likely to give some short response or ignore the robot, responds with “If you have any other questions, please let me know” and the robot returns to the service counter.

### VI. CONCLUSION

In this work we have presented a curiosity-based system for generating interactive behavior for a social robot. Our curious robot first learns socially-appropriate behavior by imitation from offline data, then continues to adapt to customers’ reactions in real time in order to satisfy its curiosity about the customers’ individual differences. We then present a case study, demonstrating an example where the robot differs its action with an engaged and unengaged customer.

We maintain that crowdsourcing of interaction data and data-driven methods of developing HRI behaviors present a promising direction for building robot behaviors in the future, but for long-term interactions, robots will also need to personalize their behavior to the user’s needs. The proposed curiosity technique illustrates one approach to this problem. However, we expect that curiosity and social appropriateness alone will not be sufficient for long-term interaction with users.

Robots driven by curiosity alone might exhibit a variety of behaviors as they explored the ways users would react to their actions, but over time we would want a robot to explore less and settle on some desirable, personalized behavior. For this, some kind of objective function based on feedback from that person is necessary – a companion robot might want to make a person happy, whereas a physical training robot might have a person’s fitness as its overall goal. In either case, curiosity could be a useful mechanism for exploring a person’s preferences and personalizing behavior. Then, once the robot’s curiosity is satisfied, the main objective function could dictate the robot’s long-term behavior in a personalized way. Thus, we believe our proposed technique constitutes an important advancement towards long-term personalization for social robots.

## VII. REFERENCES

- [1] M. Doering, P. Liu, D. F. Glas, T. Kanda, D. Kulič, and H. Ishiguro, "Curiosity did not kill the robot: A curiosity-based learning system for a shopkeeper robot," *Trans. Hum.-Robot Interact.*, In Submission.
- [2] P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro, "Data-Driven HRI: Learning Social Behaviors by Example from Human-Human Interaction," *IEEE Trans. Robot.*, vol. 32, no. 4, pp. 988–1008, 2016.
- [3] C. Breazeal, N. DePalma, J. Orkin, S. Chernova, and M. Jung, "Crowdsourcing Human-Robot Interaction: New Methods and System Evaluation in a Public Environment," *J. Hum.-Robot Interact.*, vol. 2, no. 1, pp. 82–111, 2013.
- [4] R. P. Petrick and M. E. Foster, "What would you like to drink? Recognising and planning with social states in a robot bartender domain," in *Proceedings of the 8th International Conference on Cognitive Robotics (CogRob 2012)*, (July 2012), 2012.
- [5] I. Leite, A. Pereira, A. Funkhouser, B. Li, and J. F. Lehman, "Semi-situated learning of verbal and nonverbal content for repeated human-robot interaction," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 13–20.
- [6] H. Admoni and B. Scassellati, "Data-driven model of nonverbal behavior for socially assistive human-robot interactions," in *Proceedings of the 16th international conference on multimodal interaction*, 2014, pp. 196–199.
- [7] J. E. Young, E. Sharlin, and T. Igarashi, "Teaching robots style: designing and evaluating style-by-demonstration for interactive robotic locomotion," *Human-Computer Interact.*, vol. 28, no. 5, pp. 379–416, 2013.
- [8] P. Shah, D. Hakkani-Tur, B. Liu, and G. Tur, "Bootstrapping a Neural Conversational Agent with Dialogue Self-Play, Crowdsourcing and On-Line Reinforcement Learning," *Proc. 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Vol. 3 Ind. Pap.*, pp. 41–51, 2018.
- [9] Y. Jiang, J. K. Kummerfeld, and W. S. Lasecki, "Understanding Task Design Trade-offs in Crowdsourced Paraphrase Collection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2017, pp. 103–109.
- [10] Y. Kang, J. K. Kummerfeld, M. A. Laurenzano, and A. Arbor, "Data Collection for a Production Dialogue System: A Clinic Perspective," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*, 2018, pp. 33–40.
- [11] R. Toris, D. Kent, and S. Chernova, "The Robot Management System: A Framework for Conducting Human-Robot Interaction Studies Through Crowdsourcing," *J. Hum.-Robot Interact.*, vol. 3, no. 2, pp. 25–49, 2014.
- [12] Y. Nagai, C. Muhl, and K. J. Rohlfing, "Toward designing a robot that learns actions from parental demonstrations," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 3545–3550.
- [13] M. E. Foster, S. Keizer, Z. Wang, and O. Lemon, "Machine learning of social states and skills for multi-party human-robot interaction," in *Proceedings of the workshop on Machine Learning for Interactive Systems (MLIS 2012)*, 2012, p. 9.
- [14] M. F. McTear, "User modelling for adaptive computer systems: a survey of recent developments," *Artif. Intell. Rev.*, vol. 7, no. 3–4, pp. 157–184, 1993.
- [15] G. Gordon *et al.*, "Affective Personalization of a Social Robot Tutor for Children's Second Language Skills," in *AAAI*, 2016, pp. 3951–3957.
- [16] R. Parundekar and K. Oguchi, "Learning driver preferences of pois using a semantic web knowledge system," in *Extended Semantic Web Conference*, 2012, pp. 703–717.
- [17] D. Bršćic, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-D range sensors," *IEEE Trans. Hum.-Mach. Syst.*, vol. 43, no. 6, pp. 522–534, 2013.
- [18] P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro, "Learning proactive behavior for interactive social robots," *Auton. Robots*, vol. 42, no. 5, pp. 1067–1085, Jun. 2018.
- [19] M. T. Chan, R. Gorbet, P. Beesley, and D. Kulič, "Curiosity-Based Learning Algorithm for Distributed Interactive Sculptural Systems," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, 2015, pp. 3435–3441.
- [20] Q. A. Wang, "Probability distribution and entropy as a measure of uncertainty," *J. Phys. Math. Theor.*, vol. 41, no. 6, p. 065004, 2008.
- [21] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *J. Artif. Intell. Res.*, vol. 4, pp. 129–145, 1996.
- [22] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37*, 2015, pp. 448–456.