Multimodal Sensing of Goals and Activities During Interactions With a Co-created Robot

Paras Sharma[0009−0007−7645−653X], Veronica Bella, Angela E.B. Stewart, and Erin Walker

University of Pittsburgh, Pittsburgh PA 15216, USA
{pas252,veb20,angelas,eawalker}@pitt.edu

Abstract. Culturally responsive computing (CRC) curricula engage learners in reflections on power and identity as they build technologies. Open-design tasks, with learner-chosen goals and multiple pathways to achieving them, are common in CRC and could be enhanced using adaptive technologies. Current adaptive technologies function best in well-defined learning trajectories. However, it is unclear how to design these technologies to respond to individual learners' ideas in open-design settings. In this paper, we prototype a learning system that uses multimodal sensing, log data, and reflective dialogues to build explanatory learner models in open-design settings. We deploy our system in a 2-week summer camp with middle school girls and evaluate the system's effectiveness to understand learner goals and activities. We show the importance of multiple modalities in making inferences about learner goals and activities.

Keywords: Multimodal Sensing · Open-ended Learning Systems · Culturally Relevant Computing · Goal Modeling.

1 Introduction

Open-ended learning environments (OELEs) offer a technology-enhanced learning context and scaffolding to aid students in exploring and developing solutions to real-world, complex problems [1]. We define **open-design environments** as a type of OELE where the inherent nature of the students' task makes the learning goals ambiguous and ill-defined, giving learners significant freedom to determine their goals and activities. For instance, in our study learners were tasked with "building a robot protege." Some focused on its physical design, while others programmed its functionality. Even when learners had similar goals, activities varied: some added sensors before programming the robot, while others used a mix-and-match method. This flexibility in the goals and activities enables learners to incorporate their experiences into technology design, making these environments ideal for Culturally Responsive Computing (CRC) [6] contexts, which engage learners in reflections on their design choices and how these choices relate to their identities and experiences. This reflection helps learners understand the significance of their goals, application of their learning to broader aspects of life, and consider issues of power and identity. However, the system must comprehend

learner goals and activities to prompt such reflections in open-design contexts and aid in effectively guiding learners through the learning process.

Intelligent learning technologies are powerful in learner support in well-defined trajectories (e.g. problem-solving domains) or open-ended environments with multiple pathways for constrained end goals (e.g. learning mechanical physics by interacting with real-world simulations [4]). Previous research has used text, audio, video, and log data employing methods like computational linguistics [7] and computer vision [5] to analyze learner engagement and concept development. However, these analytical techniques are best suited for well-defined domains and, on their own, cannot predict the diverse learner pathways in open-design domains. To support learning in these domains, educational systems should enable learners to define their learning outcomes, facilitate inquiry learning towards achieving those outcomes [3], and actively employ scaffolding [2] to guide learners. Per our understanding, there is a lack of systems that can realize these open-design contexts, and there is not enough information about what kinds of sensing modalities are important to build models of learner goals and activities.

In this paper, we present the design of a *multimodal sensing system* and apply it to robotics education, with the end goal of building systems to realize open-design learning. We evaluate the utility of our system to build inferences about learner goals and activities and present the significance of each modality in making these inferences, answering 2 research questions: **RQ1**: How do we elicit learner goals and activities in an open-design environment? and **RQ2**: What can we infer about the relationship between learner goals and activities?

2 System and Study Design

Our system enables learners to design the aesthetics, add physical sensors, program different functionalities, and have video or dialogue interactions with their robot. This multimodal design is driven by 3 interfaces of our system: the Frontend, the Backend, and the Hummingbird robot. The Frontend is the main interaction interface for a learner and captures their block programming and video recording activities. The Backend executes code requests from the Frontend, logs every user action, and manages the dialogue system. The dialogue system triggers 3 different interaction scenarios with the learners serving goal-identification (robot asking learners about their goals), reflection (robot discussing design choices with learners), and rapport building (robot interacting socially with learners). The Hummingbird[1] robot, an off-the-shelf robotics kit comprising a micro bit, LEDs, TriLEDs, servo motors, sound, and distance sensors, along with a custom ESP8266 Arduino board based "Sensor detection circuit" detects when a sensor/actuator is added or removed. Additionally, we equipped the robot with a camera to capture video inputs and a speaker for spoken responses.

We evaluate our system within the context of a two-week-long summer camp with 14 upper-elementary to middle-school (4^{th} to 7^{th} grades) girls, recruited

[1] https://www.birdbraintechnologies.com/products/hummingbird-bit-robotics-kit/

through our community partner. The community partner organization runs out-of-school programming for learners from a historically African American neighborhood in a mid-sized US city. The learners ranged from $8-12$ years old (average = 10.36, SD = 1.2) with 12 identifying as Black and 2 with no answer. 65% had prior experience with computer science and robotics through "tech clubs" and participation in prior robotics camps. 3 learners participated in a similarly themed camp we ran last year. The camp was distributed over 7 days, across 2 weeks, with 3 hours each day, including two 15 minute breaks.

The broader camp included: Culturally Responsive Computing, AI Fairness, Futuring Day, and Robot Co-Creation. For this paper, we dive into the Robot Co-Creation sessions, where our system was deployed. The learners were asked to: "**Create a robot protege to be presented on a robot runway**". The learners completed their robots in 4 one-hour sessions: aesthetic design, brainstorming and coding, sensor addition, and a final session to complete the design. Our study underwent ethical evaluation by our IRB, and we obtained parental consent and learner assent before starting. Moreover, learners' verbal assent was re-taken every time they were audio or video recorded during our sessions.

3 Results

We gathered 4 types of data: sensor, logs, dialogues, and video interaction data. Given our study's small sample size, we address our research questions with a descriptive analysis of the learners' actions in the different data modalities.

3.1 How do we elicit learner goals and activities in an open-design environment?

Learner Goals. We gathered learner goals via goal-eliciting dialogues, learner-initiated video recordings, and facilitator-initiated interactions. 8 out of 14 learners specified their goals with the robot (e.g. "Spin around"), with only 3 learners expressing goals related to the aesthetic design of their robot (e.g. "Add braids to my robot") using the dialogue and video modalities. 13 learners did express design goals during their interactions with the facilitators but did not explicitly state those goals during interactions with the system. The first author followed an inductive coding approach by going through all the identified goals and categorizing them based on the robot actions they indicate. After this step, we identified 5 categories representing all the learner goals: Movement (e.g. "move forward"), Light Up (e.g. "turn on LED lights"), Speech (e.g. "say it's nice to meet you"), Physical Features (e.g. "add a tail"), and other actions (e.g. "perform $8 * 5$"), with the Movement being the most common goal category across learners (8 learners had a movement goal). Learners often specified combinations of goals with most of the goals being *atomic*, which can be performed with a fixed number of system actions, (e.g. "Light Red LED"). Some learners also expressed abstract goals (e.g. making their robot "dance"), which could have different interpretations across different learners. Figure 1 shows all the goal categories.

Learner Activities. 7 learners interacted with the sensors (Mean = 2.85, SD = 1.06) with the rotation motor being the most used (Mean = 1.42, SD = 0.53), positively correlating with the high number of movement goals present among the learners. 12 learners interacted with the blocks (Mean = 4, SD = 2.41). While movement goals were the most reported, the speech block (to make the robot speak) was the most used (Mean = 0.91, SD = 0.51). On average, the longest dialogues between the robot and the learners had 7.5 words (SD = 5.9) except for 3 learners with a maximum dialogue length of 22 words. The dialogues captured learners' intent and were the main interaction source with the robot for a learner. Some learners talked about their daily activities with their robots (e.g. "play with my friends and a sleepover"), while others expressed their design goals (e.g. "turn on your LED lights from your eyes"). 10 learners used the video modality to describe their goals and activities with the robot (Mean = 5.6, SD = 4.14). Video interactions also happened during facilitator interviews, with learners using the video interface to respond. Our system did not automatically capture aesthetic design activities (actions to complete design goals).

Only 6 learners interacted with all the modalities, with dialogue (12) and blocks (12) being the most common. All the learners who added code blocks (12) had dialogues with the robot. All the learners who added sensors (7) also added code blocks and had dialogues with the robot. One of the learners just added code blocks and did dialogues with the robot but did not add any sensors.

3.2 What can we infer about the relationship between learner goals and activities?

Figure 1 shows an aggregated goal tree visualizing the connection between different sensed modalities and learner goals built using data collected by our system. A solid edge (Edge 1 in Fig. 1) positively reinforces the system's understanding of the learner goal and the sensed modality it is associated with. A dashed edge from a sensor/block node to a goal node (Edge 3 in Fig. 1) reflects the scenarios where learner goals are known by the system, however, there are no code blocks/sensors used by the learner to realize those goals. For example, 3 learners who depicted "Movement" goals did not add rotation servos to realize those goals. Another visible scenario is the absence of goals detected by the system but the additional presence of sensors/code blocks indicating the change in learner goals as they continue with their design (Edge 2 in Fig. 1). For example, 2 learners added "Light LED" and "Light Tri-LED" coding blocks but didn't express the "Light Up" goals in any of the modalities. The tree also shows no node linked to the "physical features" goal category, corresponding to aesthetic design goals. Our system could build this perception automatically during the learning interactions based on the continuous sensing of learner goals and activities. It could then take action based on what is sensed – for example, triggering reflective dialogues when goals and activities are well-understood, help-giving dialogues when there is a discrepancy between activities and goals, or goal-elicitation dialogues when the system believes it does not fully understand learner goals.

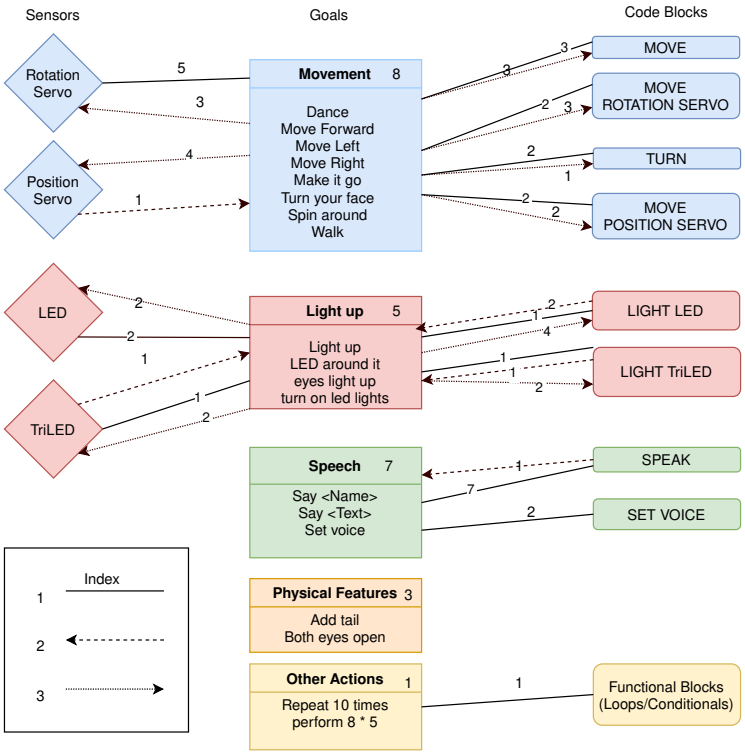

Fig. 1. Aggregated Goal Tree for 8 learners. The solid edge is the presence of both nodes together for a learner. The dashed edge from a goal node to a sensor/block node is when a sensor/block is present but the goal is not. The dashed edge from a sensor/block node to a goal node is when a goal is present but the corresponding sensor/block is not. The weight on an edge is the number of learners this pattern happened for and the weight of a goal node is the learners that had goals under this category.

4 Discussion & Conclusion

We prototype a multimodal robotics learning system for open-design environments to capture learner goals and activities. We test our system in a 2 week-long summer camp and collect learner goal and interaction data within different modalities. We then show the relationships between the sensed goals and actions.

Despite using video and dialogues, our system only captured goals for 8 learners. Since the videos were only learner-initiated, contrary to the previous works where the continuous recording of learner activities is done, we expected fewer goal-eliciting interactions through them. We had hoped that the dialogue interactions would be the main source for gathering learner goals. However, very few learners specified their goals in the dialogues. Moreover, our system typically conducted goal-eliciting dialogues with learners at the start of sessions, when they might not have planned their robot's aesthetic design and therefore did not

report such goals. Hence, we recommend future systems initiate context-based dialogues to elicit evolving goals throughout the open-design learning process. Although the aesthetic design was a significant aspect of the learner activities, it was not explicitly captured by any modality in our system. We anticipated that learner-initiated video interactions would capture these design activities. However, learners seldom showed their robots to the camera despite repeated reminders from the facilitators. This presents the need to build new data collection methods that do not violate learner privacy by continuously video recording them, as done in previous studies, and can still efficiently capture the designs.

This paper presents a system for multimodal data collection in open-design robotics learning indicating the importance of various learner interaction modalities for further exploration and application to different learning contexts. We acknowledge our small sample size but argue that it is important to do this kind of work on locally run community programs, rather than solely using large-scale online data, exploring how these methods can be applied to hyper-specific contexts that may be relatively unique. By contributing in this direction, we aim to enhance the AI-based learning environments that could support co-creation experiences in open-design learning using multimodal data analytics.

Acknowledgments. We thank Dr. Tara Nkrumah, Dr. Amy Ogan, and Dr. Kimberly Scott for leadership, and team members Angeline Pho, Paige Branagan, Nicole Balay, and Adeola Adekoya. This work is funded by NSF DRL-1811086 and DRL-1935801.

References

1. Basu, S., Biswas, G., Kinnebrew, J.S.: Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. User Modeling and User-Adapted Interaction **27**(1), 5–53 (Mar 2017)
2. Biswas, G., Segedy, J.R., Kinnebrew, J.S.: Smart open-ended learning environments that support learners cognitive and metacognitive processes. In: Holzinger, A., Pasi, G. (eds.) Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. pp. 303–310. Springer, Berlin, Heidelberg (2013)
3. Fournier-Viger, P., Nkambou, R., Nguifo, E.M.: Building Intelligent Tutoring Systems for Ill-Defined Domains, pp. 81–101. Springer Berlin Heidelberg, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14363-2_5,
4. Land, S.M., Hannafin, M.J.: Patterns of understanding with open-ended learning environments: A qualitative study. Educational Technology Research and Development **45**(2), 47–73 (Jun 1997). https://doi.org/10.1007/BF02299524
5. Raca, M., Tormey, R., Dillenbourg, P.: Sleepers' lag - study on motion and attention. In: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. p. 36–43. LAK '14, Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2567574.2567581
6. Scott, K., Sheridan, K., Clark, K.: Culturally responsive computing: a theory revisited. Learning, Media and Technology **40**, 1–25 (12 2014). https://doi.org/10.1080/17439884.2014.924966
7. Sherin, B.: A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. Journal of the Learning Sciences **22**(4), 600–638 (2013). https://doi.org/10.1080/10508406.2013.836654