Assignment 2: Paper Reviews

University of Pittsburgh, CS 2100 Research Topics in Computer Science

2

Fall 2019

1 Saerbeck, Martin, et al. "Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor." Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2010.

1.1 Technical Summary

This paper assessed the impact of a teaching robot's socially supportive behaviors on learning outcomes. 16 students (9 female, 7 male) age 10-11 were given a language learning task and each was assisted by an iCat robot. Half of the students were given a robot that displayed socially supportive behaviors in the form of role modelling, non-verbal feedback, attention guiding, empathy, and communicativeness. The other half were given a neutral iCat that did not display these behaviors.

The iCat's behavior is determined by a state-based model with 109 states, some of which are triggered by the completion of a previous state and some of which are triggered by action of the user. The iCat's input from the students involves three technologies: speech parsing, facial recognition, and visual recognition of the educational cards that the students use in their lesson. The iCat analyzed the cards to determine if the student had the right or wrong answer.

1.2 Description of Contributions

The authors propose that a child interacting with a social/emotional robotic tutor will learn more than children who interact with a non-social robotic tutor. This builds on existing literature about social robots in non-educational contexts as well as literature about social interactions among students and between students and human tutors or teachers. The authors cite previous research on animated pedagogical agents, some of which display social behaviors such as politeness. The results from the cited studies were mixed. The experiment described here involves a robot rather than an animated agent, which has the additional advantage of physical presence. The authors argue that a robot will have a stronger ability to form a social bond with students and therefore be a stronger learning partner.

The experiment validated the authors' hypothesis; results indicate that the social iCat had a positive impact on learning. The students that used the social robot showed greater improvement between the pre- and post-tests. Students with the social robot placed themselves physically closer to the robot than students with the neutral robot, which correlates with a stronger social bond. The students' answers in a post-lesson questionnaire also indicated that those with the social robot were more highly motivated to learn than those with the neutral robot.

1.3 Major Critiques

Strengths

- The paper does a nice job of clearly laying out previous research in various disciplines and how the described experiment builds on that research in a novel way.
- The authors acknowledge that there is ambiguity about what "social robot" or "social agent" means. Some studies consider an agent to be "social" merely for appearing lifelike and/or communicating verbally. To address this ambiguity, the authors modeled the robot's social interactions after a peer-reviewed article describing social interactions among students.
- The paper is well written and easy to follow.
- The study considered the gender of the students and was fairly balanced in terms of gender.

Weaknesses

- The study was on the small side with only 16 students.
- The paper mentions that "opponents criticize that animated agents might raise too high expectations..." (p1614). The iCat robot looks like a cartoon cat and the students might expect a cartoon cat to behave socially. It's possible that a cartoon cat robot that does not behave socially is actually disappointing or demotivational to the student. The study results may have been different if it involved a less cartoony robot or if it included a control group that conducted the learning task with no robot at all.
- The authors state that they believe social robots will have a higher educational impact than animated social agents, but they do not test this point in the experiment.

Questions

- I would have liked more information on system architecture and the models used for this study. For instance, how was facial recognition used? What applications were used for speech parsing, facial recognition, and card recognition?
- When the paper described how the social iCat robot interacted differently than the neutral iCat, I found myself wondering if the social iCat's utterances and interventions were of longer time duration than the neutral iCat. It's possible that students benefit from more interaction with a robot whether or not the robot displays social behaviors.
- The learning task involved students learning Toki Pona, a fake language. Did the students realize that they were learning a skill with no practical application? How does a non-practical learning task impact the students' motivation? It's possible that students in a more realistic learning environment may have reacted differently to the robots.

1.4 Minor Points

I would have appreciated more concrete examples of how a student interacted with the social iCat versus a similar interaction with a neutral iCat.

1.5 Concluding Remarks

This was a well-designed and well done study that provides support to the hypothesis that robots and intelligent agents can improve learning outcomes by interacting socially with students. The experiment's close adherence to educational literature about social interactions between humans in an educational context lends weight and legitimacy to the results and provides a framework that future studies can use to model social behaviors. The authors' use of an intelligent pedagogical robot was novel at the time of publication.

2 Kanda, Takayuki, Michihiro Shimada, and Satoshi Koizumi. "Children learning with a social robot." 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2012.

2.1 Technical Summary

This study involved giving 31 6th-grade students a learning task guided by a robot. There was no human teacher in this class but the robot was controlled remotely by a human. The robot, Robovie, displayed social behaviors to only half of the student groups. There were eight two-hour lessons on building a Lego Mindstorm robot.

The Robovie robot used cameras and microphones to convey information to the human operator. The operator followed a rules to select motion for Robovie as well as pre-recorded utterances.

2.2 Description of Contributions

The stated goal of the authors was to observe the impact of a social robot teaching a number of students rather than just one. This is a novel experiment, though technical limitations necessitated that the robot was not autonomous - it was controlled remotely by a human following a specific set of rules. An additional novel aspect of the study is the duration of the interaction: the children met with the Robovie robot for seven classes (followed by an eighth class that did not involve Robovie) while previous studies typically involved only one session with a robot.

The results suggested that the students did learn effectively from Robovie. Students with whom Robovie interacted socially did build a stronger relationship with it. Robovie's social behaviors did appear to improve learning in the first two lessons but not overall, suggesting that students benefitted from a “novelty effect” of a social robot but the effect faded over time.

2.3 Major Critiques

Strengths

- Unlike many other studies investigating student interactions with robots, this study took place over a number of class sessions and recorded how the students’ attitudes and interactions with the robots changed over time. Shorter-term studies could display results associated with a ”novelty factor” of the robot and not realistically reflect how students would interact with robots over the long term.
- This experiment involved one robot interacting with many students, which may be more realistic than one-on-one student/robot interactions.

Weaknesses

- This study was conducted ”Wizard of Oz” style, meaning that the robot was controlled remotely by a human. This raises questions about the validity of the results as an assessment of human-computer interaction.
- The group of students was very unbalanced in terms of gender - 25 boys and six girls. The results could be skewed by this disparity, particularly given the competitive nature of the teaching style described. Girls tend to learn better in a cooperative environment rather than competitive.
- The authors state that, if a student attempted to initiate an interaction with the Robovie robot (as opposed to Robovie initiating the interaction), Robovie would ignore the child. This behavior could be interpreted as rude or antisocial or it might simply decrease the robot’s social presence and decrease the impact of the robot’s social behaviors.

Questions

- One of the planned behaviors of the teaching robot, either in its social or neutral state, was that it provided only a limited amount of guidance to students (the *learner-centered approach* described in the paper). The students were meant to figure out much of the task on their own. This raises the question - did the presence of a robot help at all? If the students had conducted the Lego Mindstorm tasks without the Robovie robot and instead been shown the same videos that Robovie showed and provided with a handout with the information that Robovie provided, would they have done as well?
- The dialogue that the paper describes Robovie using may be culturally inspired (this study took place in Japan). It’s hard to imagine an American teacher saying, ”You are the first to finish your robot. Great!” This comment implicitly criticizes the students who were not the first to finish their robot. To what extent can results of any study in social robotics be extended across cultures?
- Based on the language in the paper, it’s not clear to me whether multiple groups of students participated in the same class. This is a really key point because it’s the difference between Robovie behaving socially to some classes and not others as opposed to Robovie behaving socially to some students in a class and not *other students in the same class*. A native speaker of English would likely describe the set of students who were in the classroom at the same time as a ”class” and reserve the word ”group” for a smaller subset - it’s not unusual in the U.S. to break an entire class into smaller groups of three or four to perform certain tasks. I think that the ”groups” described in the paper are small classes, but I’m not sure. If multiple groups of students, some of who interacted with Robovie in the social condition and some in the neutral condition, were combined into the same class, I would have several additional criticisms of the study.

2.4 Minor Points

The word choice and sentence structure made this paper a bit difficult to read at times, though that’s almost definitely because the writers are native speakers of Japanese.

2.5 Concluding Remarks

I think that the most important contribution made by this paper is the observation that the social behaviors of the robot seemed to impact the students only for the first two lessons, raising an often-missed point about the novelty of educational robots to children. Further study is needed to examine this novelty effect. The authors’ observations with respect to their hypotheses were modest. Overall, the social robot did not improve learning, though the students with the social robot did seem to form a stronger bond with it than the students with the neutral robot. Even these modest observations are weakened by the fact that the robot was controlled by a human. The authors attempted to move forward the state of the art by using a single robot to teach an entire class, but current technology makes the desired interactions challenging to create in an autonomous robot. Therefore, while I consider this study to have value, I do not consider it to be state of the art.

3 Gordon, Goren, et al. "Affective personalization of a social robot tutor for children's second language skills." Thirtieth AAAI Conference on Artificial Intelligence. 2016.

3.1 Technical Summary

The authors conducted a study in which preschool students, accompanied by a social robot, played games on a tablet designed to improve second language skills. The robot used facial analysis software to decide on an affective state based on the student's affective state. For half the students, the robot used a static policy to determine affect. For the other half, the robot used reinforcement learning to customize a policy with the particular student over several learning sessions. The authors hypothesized that the customized policy would result in better learning outcomes.

The techological components used for each student interaction included three items: a Tega robot that acted as a tutor for the student, a tablet on which the student conducted educational activities, and a smartphone that captured the student's facial expressions. The behavioral model for the robot used information from the tablet to determine how the student fared on the educational task and Affdex SDK to analyze the student's facial expression.

3.2 Description of Contributions

While numerous previous studies used rule-based models to generate social behaviors in a robot or intelligent agent, this study attempts to use reinforcement learning to personalize the response of the educational robot for each individual student. This paper cites previous literature studying the impacts of educational robots using reinforcement learning and personalization to model the student's cognitive state and skills, but not literature that seeks to personalize a robot's affective state to a particular student in order to maximize learning.

The study was done with preschool children aged 3-5, which is unusually young for this kind of study. The study contributed to our understanding of how pedagogical robots interact with such young kids. The results were mixed, partially due to limitations of the preschool learning environment, but indicate that affective personalization could be beneficial.

3.3 Major Critiques

Strengths

- This paper does a great job of clearly explaining the technological basis of the study without clogging up the pages with too much detail.
- The experiment takes place over two months, allowing for numerous interactions between each student and the robot. Each student had between three and seven sessions with the robot. This allows for the possibility of seeing how each student's relationship with and attitude toward the robot changes over time.
- The study did track students by gender and was somewhat balanced with 10 boys and 8 girls.

Weaknesses

- Only 18 students ended up participating in the study, which is rather small.
- While the idea of conducting the experiment with preschool children is interesting, the authors' results were impaired by limitations of the attention spans, language skills, and ability to follow directions that are fairly typical for such young children. Perhaps their next study could be designed with those limitations in mind.
- While the students all had numerous interactions with the robot, each interaction lasted only a few minutes. This is likely due to the attention span of a preschooler, but it limits the conclusions that can be drawn from the experiment.

Questions

- I wonder whether the study would have been more effective if the Tega robot were able to use a camera to capture the students' facial expressions rather than using a separate smart phone. The authors state that this was done because of technical limitations of the facial analysis software, which is understandable.
- Though the study takes place over two months, the authors don't talk about how the students attitude toward the robot changes over time. I mentioned the duration of the study in the "Strengths" section, but I would love to know more information about the time factor.
- Is it a good thing or bad thing that the reinforcement learning model did not converge? It's understandable that the students' preferences in terms of the robot's affective state changed over time; that's just how humans behave. But would a better-designed model take this into account?

3.4 Minor Points

The study took place in the same classroom as the rest of the preschool class. A divider was used to block the study participants visually from the rest of the class and each student wore headphones to block out noise. That said, it would have been ideal for the study to take place in a less distracting environment.

3.5 Concluding Remarks

For the robot that used reinforcement learning, the study did show that the robot personalized it's affective state to each student, though the reinforcement learning model did not converge. The students working with the reinforcement learning robots increased their valance (positivity) toward the robot over the course of the sessions, while the students working with the static robots had constant valance. The results did not indicate that the students with the personalizing robots learned more. The modest results indicate that this is an area worth more study. The study's use of reinforcement learning to learn from facial analysis and educational tasks is state-of-the-art and a better designed or more extensive study could reveal more definitive results.

4 Lubold, Nichola, Erin Walker, and Heather Pon-Barry. "Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion." 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2016.

4.1 Technical Summary

This paper explores the impacts of a social robot that mirrors the vocal pitch of its user in an educational context. The study involved 48 undergraduate students who were each instructed to teach a robot named Quinn to do several math problems. Each student interacted with one of three versions of Quinn: a social version that did not mirror pitch, a social version that did mirror pitch, and a control version that neither behaved socially nor mirrored pitch. The authors hypothesized that the results would vary significantly by gender and so half of the participants were male and half female.

The authors used existing standards and software, including Web Speech API, Microsoft Speech API, and Praat to allow the Quinn robot to entrain to the user's pitch.

4.2 Description of Contributions

The authors build on previous literature indicating that users who interact with spoken dialogue systems that use acoustic-prosodic entrainment have a more positive relationship with the intelligent agent than users who interact with a non-entraining system. The focus in this study is a pedagogical robot and the study is specifically attempting to use acoustic-prosodic entrainment to build rapport and increase educational outcomes. Since the authors' previous work indicates that pitch is the most important acoustic-prosodic feature that humans use to build rapport, the robot in this study entrains to the pitch of its user.

The authors were surprised by many of the results and hypothesized that limits of the experiment may have prevented some of the expected outcomes. The authors did not find significant learning gains due to the addition of either social behavior or pitch. The version of Quinn that mirrored pitch did not lead to increased rapport with the student nor did the student assign greater social presence. Some of the results suggested that the vocal pitch entrainment, combined with more effective social behavior, could be beneficial.

4.3 Major Critiques

Strengths

- Based on the previous literature cited, acoustic-prosodic entrainmnt seems to be a promising strategy to improve a pedagogical agent. The authors' hypotheses are well supported and well researched.
- The strategies used in the experiment were evidence-based. The social behaviors of the robot was based on a previous study that successfully showed improved rapport and learning success by following a specific interaction strategy. The acoustic-prosodic entrainment utilized established software packages.
- The authors provided an appropriate level of detail about the voice adaptation module of the Quinn robot. This allows the reader to understand what the robot is doing without getting bogged down in detail.
- The study was balanced in terms of gender and the authors did track gender in their data.

Weaknesses

- The number of hypotheses in this study seem crowded, which led the study to break the participants into too many groups. There were three conditions for the robot (neutral, social, and social + voice adaptive) and each of those conditions was broken into male and female groups. While the total number of participants, 48, was a decently large number for a preliminary study, each group consisted of only 8 participants.
- Given the results, it sounds like the social behaviors of the robot in the social condition may not have been effective. The authors may have done better to first establish their hypothesis with respect to the social condition and only then adding the voice adaptive version of the robot.
- The math skills that the students were learning with Quinn seem to be too easy. Five of the original 48 participants, more than 10%, were excluded from the results because they got a 100% on the pre-test and therefore had no possibility of a learning improvement. 23 of the remaining participants, 48% of the original 48, scored 100% on the post-test. Their results were still usable, but the fact that the students hit the ceiling of measurable learning gains reduces the amount of information that can be gleaned from comparing the pre- and post-test scores across groups. The authors may have learned more by making the math skills more challenging for the students.

Questions

- The authors provided very little technical detail on how the Quinn robot's social responsiveness module behaved. Considering that this module seemed to be problematic in the study, it would have been helpful to have some understanding of where they might have gone wrong.

4.4 Minor Points

I did not find any minor points to include.

4.5 Concluding Remarks

This study is very promising in terms of research area and evidence-based strategy but the results were disappointing. The ambiguous and confusing nature of the results were likely at least partially due to flaws in the experiment, including a crowded set of research questions, a poorly designed social behavioral model, and insufficiently challenging math problems. The potential impact of a study like this is exciting and I hope that the authors proceed with a better designed study.

5 Ritschel, Hannes, Tobias Baur, and Elisabeth André. "Adapting a Robot's linguistic style based on socially-aware reinforcement learning." 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, 2017.

5.1 Technical Summary

This paper outlines a proposed experiment to adapt, in real time, a robot's social behaviors based on the user's social signals. Visual sensors will capture the user's gestures and posture. A Social Signal Interpretation (SSI) framework will use this information to estimate the user's level of engagement with the robot and use this information as a reward/punishment for reinforcement learning. The robot will change its level of introversion/extroversion when generating utterances in order to maximize user engagement. The task will be a discussion between the robot and user about characters in Alice's Adventures in Wonderland. The authors expect that user engagement will increase when the robot's level of introversion/extroversion matches the user's preference.

5.2 Description of Contributions

The authors propose using a sophisticated model to analyze a user's social signals, including facial, gestural, and posture information. This will be done by the Social Signal Interpretation (SSI) framework, which was developed of the University of Augsberg by some of the authors. Previous studies have used simpler models to determine user engagement or rapport, including analyzing user utterances, physical proximity to the robot, or facial expressions.

5.3 Major Critiques

Strengths

- If the authors or another party are able to make an experiment work with the described technology, it would be the state of the art.
- While the authors have not conducted the described experiment, they provide a very good level of detail about how the experiment will be conducted. They describe the specifics of hardware, software, algorithms, and strategies.
- The model described to generate more introverted versus extroverted utterances for the Reeti robot is evidence-based, using markers from published literature to incorporate introverted/extroverted markers into the language. The idea of using introversion/extroversion as an indicator of the robot's "personality" is also based on published literature.

Weaknesses

- This is a description of an idea for research rather than a research experiment. There is no evidence to analyze.
- The more extroverted responses described in the paper involved utterances of longer duration. There could be an impact on rapport and learning related to the duration of the utterances rather than the specific content or introversion/extroversion level of the utterances.

Questions

- The more extroverted version of the Reeti robot tends to use profanity. This is based on published reseearch indicating that expletives convey extroversion. However, this language could be startling coming from an educational robot. The authors do not state the intended age of participants in this experiment, but the use of profanity could be particularly distracting, confusing, or upsetting for children.

5.4 Minor Points

I wonder if the task itself, a robot describing characters from Alice in Wonderland, could lead to a lack of engagement for participants. There is no specific goal associated with this task, and a number of participants might find it to be boring. A human being might change the topic if a conversational partner seemed bored, but the Reeti robot can only change its level of introversion/extroversion.

5.5 Concluding Remarks

The ideas proposed in this paper are exciting in that they're using state-of-the-art technology to see and analyze social signals in a way that may approach the way the human beings see and analyze social signals. Rather than using simplistic models based on a single audio or visual feature, the authors propose reading in a range of the participant's social behaviors, including facial, gestural, and posture information. The excitement is dampened, however, by the fact that the authors haven't actually done what they describe. The level of detail in this paper indicates that the authors are ready to do a study (or ready to provide information for someone else to do a study) so I look forward to seeing if it can actually be done.