Toward A Companion Robot Fostering Perseverance in Math - A Pilot Study

Lujie Chen*, Henny Admoni, Artur Dubrawski

Carnegie Mellon University

* lujiec@andrew.cmu.edu

Abstract

Challenging math problems without immediate solutions often invite students to ride an "emotional roller-coaster" [2] through episodes of confusion, frustration, surprise and joy. Those problem solving experiences provide rich opportunities to cultivate *mathematical perseverance*, the mentality to forge ahead in face of ambiguity or difficulty. An ideal teacher closely monitors the problem solving process and provides cognitive, emotional or social supports that are often personalized and optimized. Given the potential high cognitive loads on the teacher who needs to monitor and react in real time, an affect sensitive social robot has the potential to assist by partnering with human teacher in a busy classroom. In this paper, we will describe a multi-modal dataset we collected from multiple sessions of a young child solving math problems coached by his parent tutor. We report initial findings and their implications in the interaction design of a robotic companion that responds dynamically to the child's fine-grained non-verbal behaviors cues and affect signals in order to foster perseverance. We also describe an ongoing study involving multiple parent-child pairs with additional data elements.

Introduction

Different from math exercises, non-routine math problems are those without immediate solutions. Regular exposures to those challenges help students to cultivate perseverance in face of uncertainty and impasses. Students at young age benefit from high quality coaching with personalized supports adaptive to students' moment-by-moment cognitive,emotional and social needs [8]. Unfortunately, this level of support is often not feasible given the large student-to-teacher ratio in regular classrooms.

Socially assistive robots have been explored recently in education to regulate timing of breaks [11], shape help seeking behaviors [12] or cultivate growth mindset [10], curiosity [5] or creativity [7]. There is not much exploration yet in the area of *perseverance* in math problem solving context among young students.

We envision a companion robot who can partner with teachers to sense and interpret children's behavioral cues in real time and decide on timing and types of support (cognitive, emotional or social) in coaching problem solving in regular classrooms. It is well known that supports offered too early will deprive students of learning opportunities from productive struggles [9], but delayed support might induce excessive frustration that will undermine children's confidence. Experienced human teachers fine tune those decision rules possibly from a large number of interactions with children via trial and error. Can a robot learn from human teachers on those critical decisions rules?

We approach this question by observing parents coaching their children in math problem solving at home, which gives us the opportunity to observe how the“home teachers” who have deep understanding of their own children’s ability and personality would respond to their behavior cues when offering supports. We believe that observing parents may give us a better chance to find an optimal response model, as opposed to observing school teachers. They work with large groups of students and are often trained to teach well-defined cognitive skills with less focus on developing perseverance in students.

In following sections, we will describe the data we collected from a pilot study on one parent-child pair and the response model of the tutor to the child’s non-verbal behaviors cues. We focus on head pose and eye gaze changes as well as affective signals such as frustration and confusion. We choose to study those cues as they appear to be used by tutors in deciding when and how to intervene. We will also describe an ongoing data collection effort from multiple parent-child pairs augmented with additional modalities and survey data.

Methods

Data Collection

In the first study, we recorded 21 videos of one-to-one problem solving sessions between a 9-year-old boy (a third grader) and his mother (the first author of this paper) as his tutor. Each session began at a time when the child was presented with a problem and ended when the child solved the problem, in some cases with the tutor’s help. The videos were captured in a home environment using a Logitech 1080P webcam with an integrated microphone. The seating position of the tutor and the child (Figure 1) makes it possible to capture the child’s overt intent-to-connect (ITCs), defined as head pose and eye gaze toward the tutor. We captured audio and video from the child, and only audio from the tutor.

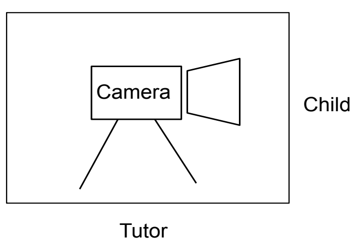

Figure 1. Recording setup. In this seating position, the child’s intent-to-connects (ITCs or head pose and eye gaze toward the tutor) are detectable. The child-facing camera captures the frontal view of the child’s upper body and face.

In a second ongoing study, we are collecting data from about 10 parent-child pairs in their home environment with the similar setup as in first study, augmented with additional data features. In addition to the videos, we also collect children’s handwriting traces during problem solving, recorded along with their speech using the Livescribe Smartpen and notebook [1]. We anticipate this additional modality will allow us to gain additional insight into children’s cognitive processes, complementing what is available from videos and audio alone. This data stream is especially valuable at times when children are working on problems silently with their heads down because in those cases neither speech nor facial expression information is available. In addition, we collect survey data using validated instruments for children’s personality traits [6], grit [4], math interests and self-regulation as well as math coaching activities at home. We also collect before and after session questionnaires on children’s emotional states.

Annotation and feature extraction

As the second study is still ongoing, the remainder of the paper focuses on the data from the first study. We use the manual annotations from all videos in the first study as the basis for our analysis. This includes the child’s ITCs and the tutor’s verbal

[1] http://www.livescribe.com/

responses. Additionally, we use the affect scores extracted from FACET [2] toolkit on the child's facial expressions, with respect to joy, surprise, frustration and confusion. Our previous study [2] shows that ITCs and tutor's verbal responses can be reliably detected using machine learning models built from features extracted from Openface [1] and COVAREP [3] respectively. We also note that FACET has reasonable accuracy when validated by human annotations from sampled video frames. As such, in our future work with the second data set, we plan to apply those validated detectors to analyze videos and audio.

Results

In the first study, we recorded a total of 21 sessions, accumulating 141 minutes of raw videos with mean length of 6.4 minutes per session, with longest session lasting 14.6 minutes and the shortest only about 2 minutes. We focus our analysis on two types of signals likely used by the tutor in response decisions: the overt non-verbal behavioral cues of ITCs and the more subtle affect signals.

Does the tutor pays attention to affect signals in her response to the child's overt ITCs?

Our analysis suggests that affect signals play a role in the tutor's response model to the child's overt ITCs. This is supported by comparing two linear regression models to explain the delay of the tutor's response using a combination of ITCs and affect features. Those models are built on 101 instances of ITCs which occurred when child looked at tutor in silence, which represents 17% of 774 ITCs across all sessions.

Model 1: $ResponseDelay = DurationOfITC + \epsilon$
Model 2: $ResponseDelay =$
$DurationOfITC + Neutral + Confusion + Frustration + Joy + Surprise + \epsilon$

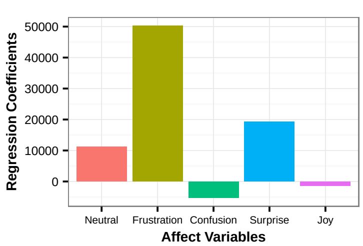

Figure 2. Regression coefficients for affect variables from Model 2. Large and positive coefficients indicate more delayed response to a specific affect comparing to others. Only coefficient for frustration is significantly different from zero.

In the first model, only the duration of ITC is taken into account, while in second model, we additionally include affect features extracted from FACET within the 5s window of each ITC. We then performed a likelihood ratio test comparing the goodness-of-fit between those two nested models with $\chi^2(5) = 12.4$ and $p = .03$. This result suggests that including affect related features improves the fitting of the second model comparing with first model. In other words, the tutor seems to take into account the contextual affect around ITCs, in addition to how long the child looks at her, in her decision of when to respond.

Figure 2 shows the relative magnitude of the regression coefficients estimated from model 2 as described above. It is interesting to note that the tutor's responses to ITCs vary by the affect context surrounding ITCs. In particular, the large positive magnitude of coefficient of frustration affect suggest observing those type of signals significantly increases her delay in response, holding all other variables constant. This seems to be counterintuitive , but it might make sense if her goal is indeed to increase the child's exposure to frustration therefore provide him with an opportunity to persevere. Confusion has a negative but non-significant coefficient, which could partly be explained by the positive correlation between frustration and confusion noted from our previous study [2]. On the other hand, joy seems to be responded to relatively faster than

[2] https://imotions.com/emotient/

surprise, which could due to the fact that most of the joy affects occurred when the child solved the problems successfully toward the end, in which cases, tutor immediately offered compliments such as "good job"; while surprise is often associated with the "aha" moment the child encountered in middle of the process, in which case, the tutor would just let him continue the discovery journey on his own.

How does tutor respond to child's standalone negative affect ?

In this section, we shift focus to a response model as related to the standalone negative affect episodes (i.e., confusion and frustration) that occurred independent of ITCs. Those episodes were observed in scenarios where the child displayed negative affect without looking at the tutor. Those affect signals are thus different from those that occurred along with ITCs, as analyzed in previous section.

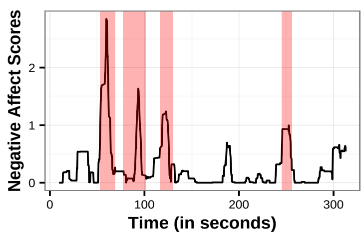

Figure 3. An example of detected negative affects (Frustration + Confusion) episodes. This plot shows the output of the automatic negative episode detector (red colored bands) overlaid with smoothed time series of negative affect for video No.12

In order to detect episodes of negative affect, we post-processed the raw output of affect scores and identified exceedances where the negative affect scores are above a certain threshold (α). we then cluster those exceedances into episodes. Two consecutive exceedances are clustered into the same episode if they occur within β seconds. we then filter out those episodes lasting less than γ seconds. We calibrate those parameters by visually inspecting the time series of affect scores overlaid with detected negative affect episodes, as shown in Figure 3 as one example. With α=.5 ,β=5s, and γ=5s we identified 92 negative affect episodes. We verified that none of the episodes overlaps with any of the non-dialogue ITCs analyzed in previous section.

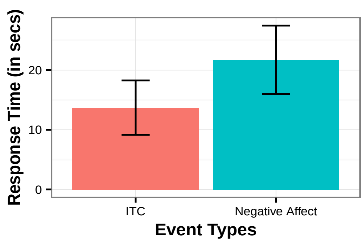

Figure 4. Comparison of response time toward standalone negative affect episodes and non-dialogue ITCs Mean of difference in response time to two different types of events, with 95% confidence intervals.

As shown in Figure 4,there was a significant difference in the response time to 92 negative affect episodes (M=21.7s, SD=2.9s) and 101 non-dialogue ITCs (M=13.7s and SD=2.3s); t(177.54)=2.14, p=0.034. Comparing with non-dialogue ITCs, the delayed response to negative affect might reflect tutor's interpretation of standalone negative affect as more definite signals of help request, in which cases, she would rather to let the child struggle longer.

Conclusion

In this paper, we analyzed the tutor's response time to the child's affect signals and non-verbal behaviors cues of ITCs. We note that tutor responded differently to those two different types of signals. we also notice that tutor seems to delay the response to negative affect whether or not they are accompanied by ITCs, which could link to tutor's intention to maximize child's opportunity to practice perseverance. We believe this line of investigation has the promise to provide objective and useful insights into the interaction design for affect sensitive companion robots in fostering perseverance in child's math problem solving activities.

Acknowledgement

The research reported here was supported in part by a training grant from the Instituteof Education Sciences (R305B150008). Opinions expressed do not represent the views of the U.S. Department of Education.

References

1. T. Baltrušaitis, P. Robinson, and L. P. Morency. OpenFace: an open source facial behavior analysis toolkit. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2016.

2. L. Chen, X. Li, Z. Xia, Z. Song, L.-P. Morency, and A. Dubrawski. Riding an emotional roller-coaster: A multimodal study of young child's math problem solving activities. In *EDM*, pages 38–45, 2016.

3. G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer. Covarep—a collaborative voice analysis repository for speech technologies. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 960–964. IEEE, 2014.

4. A. L. Duckworth, C. Peterson, M. D. Matthews, and D. R. Kelly. Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, 92(6):1087–1101, 2007.

5. G. Gordon, C. Breazeal, and S. Engel. Can children catch curiosity from a social robot? In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 91–98. ACM, 2015.

6. T. A. Judge, C. A. Higgins, C. J. Thoresen, and M. R. Barrick. The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, 52(3):621–652, 1999.

7. P. H. Kahn Jr, T. Kanda, H. Ishiguro, B. T. Gill, S. Shen, J. H. Ruckert, and H. E. Gary. Human creativity can be facilitated through interacting with a social robot. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 173–180. IEEE Press, 2016.

8. A. Kapoor, S. Mota, and R. W. Picard. Towards a Learning Companion that Recognizes Affect. *AAAI Fall symposium*, (543):2–4, 2001.

9. M. Kapur. Productive failure in learning math. *Cognitive Science*, 38(5):1008–1022, 2014.

10. H. W. Park, R. Rosenberg-Kima, M. Rosenberg, G. Gordon, and C. Breazeal. Growing growth mindset with a social robot peer. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 137–145. ACM, 2017.

11. A. Ramachandran, C.-M. Huang, and B. Scassellati. Give me a break!: Personalized timing strategies to promote learning in robot-child tutoring. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 146–155. ACM, 2017.

12. A. Ramachandran, A. Litoiu, and B. Scassellati. Shaping productive help-seeking behavior during robot-child tutoring interactions. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 247–254. IEEE Press, 2016.